

Cuk-H03680-9-P024840

**ANNUAL REVIEW OF
MEDICINE:**

Selected Topics in
the Clinical Sciences



P, 24 840

EDITORIAL COMMITTEE (1979)

C. H. COGGINS
Harvard Medical School

W. P. CREGER
Stanford University

R. M. DONALDSON, JR.
Yale University

E. W. HANCOCK
Stanford University

S. L. KATZ
Duke University

P. G. QUIE
University of Minnesota

R. M. ROSE
University of Texas

G. T. SHIRES
Cornell University

M. SIPERSTEIN
University of California Service
Veterans Administration Hospital,
San Francisco

Responsible for the organization of Volume 30
(Editorial Committee, 1977)

C. H. COGGINS
Harvard Medical School

W. P. CREGER
Stanford University

E. W. HANCOCK
Stanford University

K. J. ISSELBACHER
Massachusetts General Hospital

E. J. SACHAR
Columbia University

G. T. SHIRES
Cornell University

M. SIPERSTEIN
University of California Service
Veterans Administration Hospital,
San Francisco

R. T. SMITH
The University of Florida

L. W. WANNAMAKER
The University of Minnesota

S. A. WELLS (Guest)
Duke University

Production Editor

MARGOT PLATT

ANNUAL REVIEW OF MEDICINE:

Selected Topics in
the Clinical Sciences

WILLIAM P. CREGER, *Editor*
Stanford University School of Medicine

C. H. COGGINS, *Associate Editor*
Harvard Medical School

E. W. HANCOCK, *Associate Editor*
Stanford University School of Medicine

VOLUME 30

1979

ANNUAL REVIEWS INC.
Palo Alto, California, USA

COPYRIGHT © 1979 BY ANNUAL REVIEWS INC., PALO ALTO, CALIFORNIA, USA. ALL RIGHTS RESERVED. The appearance of the code at the bottom of the first page of an article in this serial indicates the copyright owner's consent that copies of the article may be made for personal or internal use, or for the personal or internal use of specific clients. This consent is given on the condition, however, that the copier pay the stated per-copy fee of \$1.00 per article through the Copyright Clearance Center, Inc. (P. O. Box 765, Schenectady, NY 12301) for copying beyond that permitted by Sections 107 or 108 of the US Copyright Law. The per-copy fee of \$1.00 per article also applies to the copying, under the stated conditions, of articles published in any Annual Review serial before January 1, 1978. Individual readers, and nonprofit libraries acting for them, are permitted to make a single copy of an article without charge for use in research or teaching. This consent does not extend to other kinds of copying, such as copying for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale.

REPRINTS The conspicuous number aligned in the margin with the title of each article in this volume is a key for use in ordering reprints. Available reprints are priced at the uniform rate of \$1.00 each postpaid. The minimum acceptable reprint order is 5 reprints and/or \$5.00 prepaid. A quantity discount is available.

International Standard Serial Number: 0066-4219
International Standard Book Number: 0-8243-0530-2
Library of Congress Catalog Card Number: A51-1659

Annual Reviews Inc. and the Editors of its publications assume no responsibility for the statements expressed by the contributors to this Review.

2197

PRINTED AND BOUND IN THE UNITED STATES OF AMERICA

CONTENTS

[Note: Titles of chapters in Volumes 26-30 are arranged by category on pages 508-14.]

THE ARTERIAL WALL AND ATHEROSCLEROSIS, <i>Russell Ross</i>	1
HAIRY CELL LEUKEMIA, <i>Raul C. Braylan and Jerome S. Burke</i>	17
BIOLOGICAL AND ENZYMATIC EVENTS IN CHEMICAL CARCINOGENESIS, <i>Henry C. Pitot</i>	25
ANGIOGRAPHIC MANAGEMENT OF GASTROINTESTINAL BLEEDING, <i>Saadoon Kadir and Christos A. Athanasoulis</i>	41
THE MEDICAL MANAGEMENT OF PEPTIC ULCER, <i>David J. C. Shearman and David Hetzel</i>	61
VALIDITY OF THERAPY FOR MILD HYPERTENSION, <i>Edward D. Freis</i>	81
SICK SINUS SYNDROME, <i>J. Thomas Bigger, Jr. and James A. Reiffel</i>	91
SECRETABLE STORAGE POOLS IN PLATELETS, <i>Holm Holmsen and Harvey J. Weiss</i>	119
CLONAL ORIGIN OF HUMAN TUMORS, <i>Philip J. Fialkow</i>	135
THE SYNDROME OF HYPORENINEMIC HYPOALDOSTERONISM, <i>James P. Knochel</i>	145
DIETARY MANAGEMENT OF DIABETES MELLITUS, <i>John W. Ensink and Edwin L. Bierman</i>	155
THE PROLONGED QT INTERVAL, <i>J. A. Abildskov</i>	171
COMPUTED TOMOGRAPHY OF THE HEPATOBILIARY SYSTEM, <i>Melvyn Korobkin and Henry I. Goldberg</i>	181
COMPUTED TOMOGRAPHY IN NEUROLOGIC DIAGNOSIS, <i>Anne G. Osborn</i>	189
LOW MOLECULAR WEIGHT PROTEINURIA, <i>C. L. Hall and J. Hardwicke</i>	199
THE ULTRASTRUCTURE OF GLOMERULAR FILTRATION, <i>Morris J. Karnovsky</i>	213

LOCALIZATION AND ANTIBIOTIC MANAGEMENT OF URINARY TRACT INFECTION, <i>Leslie S. T. Fang, Nina E. Tolkoﬀ-Rubin, and Robert H. Rubin</i>	225
BIG ACTH AND BRONCHOGENIC CARCINOMA, <i>Rosalyn S. Yalow</i>	241
PROLACTIN AND THE DIAGNOSIS OF PITUITARY TUMORS, <i>William B. Malarkey</i>	249
STUDIES OF ATRIAL FLUTTER FOLLOWING OPEN HEART SURGERY, <i>Albert L. Waldo, James L. Wells, Jr., Vance J. Plumb, Terry B. Cooper, and William A. H. MacLean</i>	259
CONTROL OF THE IMMUNE SYSTEM BY INHIBITOR AND INDUCER T LYMPHOCYTES, <i>H. Cantor</i>	269
PROSTATITIS, <i>Edwin M. Meares, Jr.</i>	279
REGRESSION OF ATHEROSCLEROSIS, <i>Donald W. Crawford and David H. Blankenhorn</i>	289
MANAGEMENT OF INSOMNIA, <i>Constantin R. Soldatos, Anthony Kales, and Joyce D. Kales</i>	301
DIAGNOSIS AND MEDICAL MANAGEMENT OF INSULINOMAS, <i>Stefan S. Fajans and John C. Floyd, Jr.</i>	313
VITRECTOMY SURGERY FOR DIABETIC RETINOPATHY, <i>George H. Bresnick and Franklin L. Myers</i>	331
DIABETIC KETOACIDOSIS: REAPPRAISAL OF THERAPEUTIC APPROACH, <i>Abbas E. Kitabchi, Ruth Young, Harold Sacks, and Lawrence Morris</i>	339
DISSEMINATED INTRAVASCULAR COAGULATION: A REAPPRAISAL, <i>Robert W. Colman, Stanley J. Robboy, and John D. Minna</i>	359
IMMUNOBIOLOGY OF THE MATERNAL-FETAL RELATIONSHIP, <i>Ross E. Rocklin, John L. Kitzmiller, and Michael D. Kaye</i>	375
GROWTH CONSIDERATIONS IN THE RADIATION THERAPY OF CHILDREN WITH CANCER, <i>Philip S. Littman and Giulio J. D'Angio</i>	405
PRENATAL SEX HORMONES AND THE DEVELOPING BRAIN: EFFECTS ON PSYCHOSEXUAL DIFFERENTIATION AND COGNITIVE FUNCTION, <i>Anke A. Ehrhardt and Heino F. L. Meyer-Bahlburg</i>	417
TREATMENT OF PLASMA CELL MYELOMA, <i>Daniel E. Bergsagel</i>	431

PATTERNS OF CIRCULATING COMPLEMENT IN RENAL DISEASES, <i>Edmund J. Lewis</i>	445
IMMUNIZATION AGAINST TYPHOID FEVER, <i>John W. Warren and Richard B. Hornick</i>	457
ENDOCRINOLOGY OF THE PLACENTA, <i>Royice B. Everett and Paul C. MacDonald</i>	473
THE EFFICACY OF BIOFEEDBACK THERAPY, <i>Martin T. Orne</i>	489
CUMULATIVE INDEX OF CONTRIBUTING AUTHORS, VOLUMES 26-30	505
CUMULATIVE INDEX OF CHAPTER TITLES, VOLUMES 26-30	508

ANNUAL REVIEWS INC. is a nonprofit corporation established to promote the advancement of the sciences. Beginning in 1932 with the *Annual Review of Biochemistry*, the Company has pursued as its principal function the publication of high quality, reasonably priced Annual Review volumes. The volumes are organized by Editors and Editorial Committees who invite qualified authors to contribute critical articles reviewing significant developments within each major discipline.

Annual Reviews Inc. is administered by a Board of Directors whose members serve without compensation. The Board for 1979 is constituted as follows:

Dr. J. Murray Luck, Founder and Director Emeritus of Annual Reviews Inc.

Professor Emeritus of Chemistry, Stanford University

Dr. Joshua Lederberg, President of Annual Reviews Inc.

President, The Rockefeller University

Dr. James E. Howell, Vice President of Annual Reviews Inc.

Professor of Economics, Stanford University

Dr. William O. Baker, *President, Bell Telephone Laboratories*

Dr. Robert W. Berliner, *Dean, Yale University School of Medicine*

Dr. Sidney D. Drell, *Deputy Director, Stanford Linear Accelerator Center*

Dr. Eugene Garfield, *President, Institute for Scientific Information*

Dr. William D. McElroy, *Chancellor, University of California, San Diego*

Dr. William F. Miller, *Professor of Computer Science and Business,
Stanford University*

Dr. Colin S. Pittendrigh, *Director, Hopkins Marine Station*

Dr. Esmond E. Snell, *Professor of Microbiology and Chemistry,
University of Texas, Austin*

Dr. Harriet Zuckerman, *Professor of Sociology, Columbia University*

The management of Annual Reviews Inc. is constituted as follows:

John S. McNeil, Chief Executive Officer and Secretary-Treasurer

William Kaufmann, Editor-in-Chief

Lawrence C. Russell, Business Manager

Sharon E. Hawkes, Production Manager

Ruth E. Severance, Promotion Manager

Annual Reviews are published in the following sciences: Anthropology, Astronomy and Astrophysics, Biochemistry, Biophysics and Bioengineering, Earth and Planetary Sciences, Ecology and Systematics, Energy, Entomology, Fluid Mechanics, Genetics, Materials Science, Medicine, Microbiology, Neuroscience, Nuclear and Particle Science, Pharmacology and Toxicology, Physical Chemistry, Physiology, Phytopathology, Plant Physiology, Psychology, and Sociology. The *Annual Review of Public Health* will begin publication in 1980. In addition, four special volumes have been published by Annual Reviews Inc.: *History of Entomology* (1973), *The Excitement and Fascination of Science* (1965), *The Excitement and Fascination of Science, Volume Two* (1978), and *Annual Reviews Reprints: Cell Membranes, 1975-1977* (published 1978). For the convenience of readers, a detachable order form/envelope is bound into the back of this volume.

ERRATA

Volume 29 (1978)

In "The Disposition and Effects of Sedatives and Analgesics in Liver Disease" by Anastacio M. Hoyumpa, Jr., Robert A. Branch, and Steven Schenker:

Morphine and meperidine should be deleted from Table 1 on page 208;

The sentence beginning on line 24 of page 208, "At present there are no known examples of drugs with high intrinsic free clearance and low protein binding," should be deleted.

THE ARTERIAL WALL AND ATHEROSCLEROSIS

◆7302

Russell Ross, Ph.D.

Department of Pathology SM-30, School of Medicine,
University of Washington, Seattle, WA 98195

INTRODUCTION

Although atherosclerosis has been recognized as the principal cause of death in Western man, it is only within the last decade or so that the disease process itself has begun to be understood (1). The disease develops insidiously before the development of symptoms. Its widespread occurrence at an early age was not generally accepted until studies of American males who died in the Korean and Vietnam wars documented its high prevalence within the second decade of life (2, 3).

The lesions of atherosclerosis are limited principally to the innermost layer of the artery wall, the intima. Although there are many variants of the lesion in man, there is general acceptance that the disease process consists of three events, each of which involves the cells of the artery wall. These include (a) intimal smooth muscle cell proliferation, (b) formation of large amounts of connective tissue matrix by the proliferated smooth muscle, and (c) deposition of lipids both within the cells and in the connective tissues surrounding them (4-7).

The principal biological phenomenon in the development of the lesions of atherosclerosis with which this article is concerned is intimal smooth muscle proliferation. Because of the accumulation of lipid and calcium deposits, atherosclerosis has commonly been considered to be a degenerative process. However, in the absence of intimal smooth muscle proliferation there would probably be few sequelae of clinical significance.

Epidemiologic research has identified a list of risk factors that are commonly associated with an increased incidence of atherosclerosis. Each of these risk factors must be explained at the cellular level if we are to understand the role they play in the induction of lesion formation (1, 8-13).

THE NORMAL ARTERY WALL

Arteries consist of three morphologically distinct layers: the intima or innermost layer; the middle layer, the media; and the external layer, the adventitia. The intima is bounded on the lumen by a continuous layer of endothelial cells that form a protective barrier between the blood and the artery wall. These cells control the entry of nutrients and have been shown to turn over at differing rates in different parts of the artery wall (14–16). At birth the intima is a narrow layer bounded by endothelium on its innermost aspect and externally by the internal elastic lamina, a fenestrated sheet of elastic tissue. There are relatively few cells within the intima at birth; however, it gradually thickens with progressive age as a result of the accumulation of small numbers of smooth muscle cells and connective tissue. The media consists entirely of smooth muscle cells oriented in a spiral fashion, surrounded by connective tissues, including collagen and glycosaminoglycan, as well as elastic fibers. The adventitia contains both smooth muscle cells and fibroblasts together with relatively large amounts of collagen, glycosaminoglycan, and elastic fibers. The changes associated with the lesions of atherosclerosis occur principally within the intima.

THE LESIONS OF ATHEROSCLEROSIS

Pathologists have subdivided the lesions of atherosclerosis into three general varieties. These are the fatty streak, the fibrous plaque, and the so-called complicated or advanced lesion (1, 4, 17).

The fatty streak is most commonly found in young individuals and is characterized by a focal accumulation of small numbers of smooth muscle cells within the intima, which generally contain deposits of lipid and are also surrounded by lipid extracellularly (18). Clinically it has a yellow sessile appearance and causes no obstruction or symptomatology. Virtually every child has such lesions by the age of ten (1). They are distributed randomly throughout the arterial tree, including those sites where increased shear stress has been shown to occur, namely at junctions and at bifurcations.

The fibrous plaque is considered characteristic of the advancing lesions of atherosclerosis and is not as ubiquitous as is the fatty streak. It has a grossly white appearance and is elevated, so that it may protrude into the lumen of the artery. Fibrous plaques consist of an intimal accumulation of proliferated smooth muscle cells surrounded by fairly large amounts of connective tissue, which appear to form a "cap" that covers a deeper accumulation of both intra- and extracellular lipids. Necrosis and cell debris are usually associated with this lipid accumulation. Fibrous plaques generally occur at sites where increased shear may be applied to the artery wall,

particularly in individuals who are hypertensive. Such sites may be at greater risk of endothelial injury (19).

The advanced or so-called complicated lesion appears to be a fibrous plaque that has been altered as a result of hemorrhage, calcification, necrosis, thrombosis, and further cell degenerative change mixed with cell proliferation. Calcification is perhaps the most distinctive change associated with the complicated lesion.

Thus the lesions of atherosclerosis are focal and contain newly proliferated smooth muscle cells, together with newly formed connective tissue, and deposits of lipid.

THE RESPONSE-TO-INJURY HYPOTHESIS OF ATHEROGENESIS

Together with my colleagues at the University of Washington School of Medicine, I have been examining a hypothesis termed the "response-to-injury hypothesis" (4, 20, 21). This hypothesis evolved because of the strong similarity in appearance between the early lesions of atherosclerosis, the "fibromusculoelastic lesions" (1), and the response of arteries to experimentally induced injury. It represents a modification of some of the earlier notions of the famed pathologist, Rudolph Virchow (22), which have since been modified by numerous investigators.

Simply stated, the response-to-injury hypothesis suggests that alterations occur to the lining endothelial cells, which thus changes the ability of cells to function as a protective barrier. The alterations in the endothelium may be quite subtle, in which case the changes may be manifest by altered permeability characteristics of the cells. On the other hand, the alterations may be sufficiently severe so that the endothelial cells may become detached from one another and possibly even from the underlying connective tissue and be swept into the blood stream. Platelets would be capable of adhering to the exposed subendothelial connective tissue and of undergoing the "platelet release reaction." The loss of the endothelial barrier function and the presence of platelets at the sites of exposed connective tissue would lead to the ingress of material derived from the platelet release reaction from the plasma into the artery wall. The interaction between plasma components and platelet constituents upon the smooth muscle cells of the artery wall is postulated to result in the focal migration of smooth muscle cells from the media into the intima. Together with the proliferation of preexisting intimal smooth muscle cells, this produces a fibromusculoelastic precursor-type lesion of atherosclerosis. Such focal sites of smooth muscle cell proliferation would be accompanied by the formation of relatively large amounts of connective tissue matrix. The smooth muscle cell is not only contractile

but also quite capable of producing relatively large amounts of all of the known connective tissue matrix macromolecules (23-27).

The hypothesis goes on to suggest that if the injury is a single event, the proliferative response that is evoked may be reversible and the lesions will regress. On the other hand, if the injury occurred on numerous occasions over an extended period of time, such cycles of injury, cell proliferation, and connective tissue formation could lead to a chronic progressive lesion that in time could become sufficiently large so as to cause the development of clinical sequelae.

This hypothesis has been a valuable aid to a number of investigators because it has indicated those areas in our knowledge that are lacking and has provided new directions for future research. If the hypothesis is to be adequately tested, it will be necessary to fully understand the biology of the two principal cells of the artery, the endothelium and the smooth muscle, as well as that of the platelet and the plasma. Each risk factor will have to be explained in cellular and molecular terms. The effects of stress, hypertension, hyperlipidemia, smoking, diabetes, and various genetic factors should be understandable on the basis of endothelial injury and the subsequent smooth muscle proliferative response if this hypothesis is to have continued merit. This hypothesis is illustrated and further explained in Figure 1.

EXPERIMENTAL DATA RELATED TO THE RESPONSE-TO-INJURY HYPOTHESIS

It is possible to disrupt the endothelial barrier in a number of different ways. The endothelium can be injured mechanically, chemically, immunologically, and by various toxins. Regardless of the source of injury, it can be experimentally shown that endothelial disruption leads to an immediate response of the platelets. Within just a few moments after the endothelium has been abraded with an intraarterial balloon catheter, platelets adhere to the subendothelial connective tissue at the site of injury (28, 29). After adhering they undergo aggregation and lose their granules during the release reaction. Such platelet adherence, aggregation, and release may occur for as long as 48 hours after mechanical injury and possibly intermittently for longer periods (Figure 2). Mechanical injury (30, 31) as well as several forms of chemical injury were shown by Harker and his colleagues (32, 33) to result in a decrease in platelet survival when experimental animals were given ^{51}Cr -tagged platelets. There appears to be a linear correlation between the amount of endothelium removed with an intraarterial balloon catheter and the extent of the decrease in platelet survival. In like fashion, the return to normal platelet survival correlates with the regeneration of the endothelium and restoration of the defect. Thus, platelet survival may serve as a useful clinical index of endothelial injury (48).

Examination of sites where platelets have adhered and undergone the release reaction showed that after three to seven days smooth muscle cells had migrated from the media into the intima between fenestrae of the internal elastic lamina and had begun to proliferate within the intima at these sites. In monkeys made hyperlipidemic by a high fat diet, not only does smooth muscle proliferate and new connective tissue form, but lipid deposition is found as well.

If the monkeys are normocholesterolemic, the intimal smooth muscle proliferative response appears to be reversible. In contrast, catheter-induced

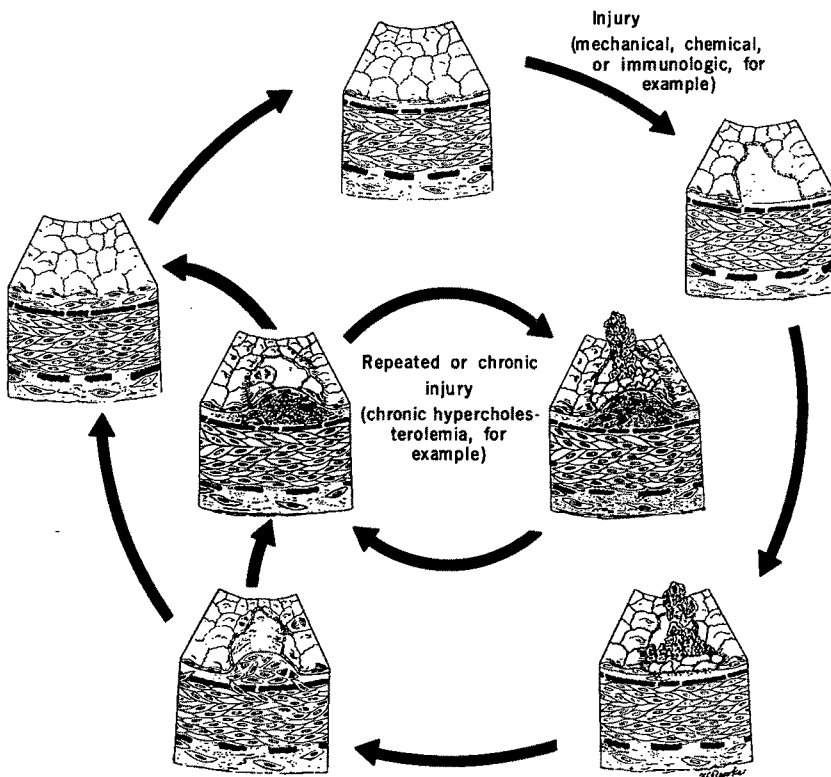


Figure 1 In the response-to-injury hypothesis, two different cyclic events may occur. The outer, or regression, cycle may represent common single occurrences in all individuals: Endothelial injury leads to desquamation, platelet adherence, aggregation, and release, followed by intimal smooth muscle proliferation and connective tissue formation. If the injury is a single event, the lesions may go on to heal and regression occur. The inner, or progression, cycle demonstrates the possible consequences of repeated or chronic endothelial injury as may occur in chronic hyperlipidemia. In this instance, lipid deposition as well as continued smooth muscle proliferation may occur after recurrent sequences of proliferation and regression, and these may lead to complicated lesions that calcify. Such lesions could go on to produce clinical sequelae such as thrombosis and infarction. Reproduced from *Science* (21).

lesions in hypercholesterolemic monkeys appear to progress and show no signs of regression (Figure 3). In addition, endothelial regeneration appears to be defective in the chronic hypercholesterolemic monkeys. Thus the sequence of events noted in the response-to-injury hypothesis can be shown to occur in animals that are injured experimentally with various forms of intraarterial mechanical devices.

A similar sequence of events can be shown to occur as a result of chemically induced injury. Such injury can occur in baboons with chronic homocysteinemia (32, 33), as a result of chronic hypercholesterolemia (21), or through immune injury that results from circulating antigen-antibody complexes (34, 35). In both chronic homocysteinemia and chronic hypercholes-

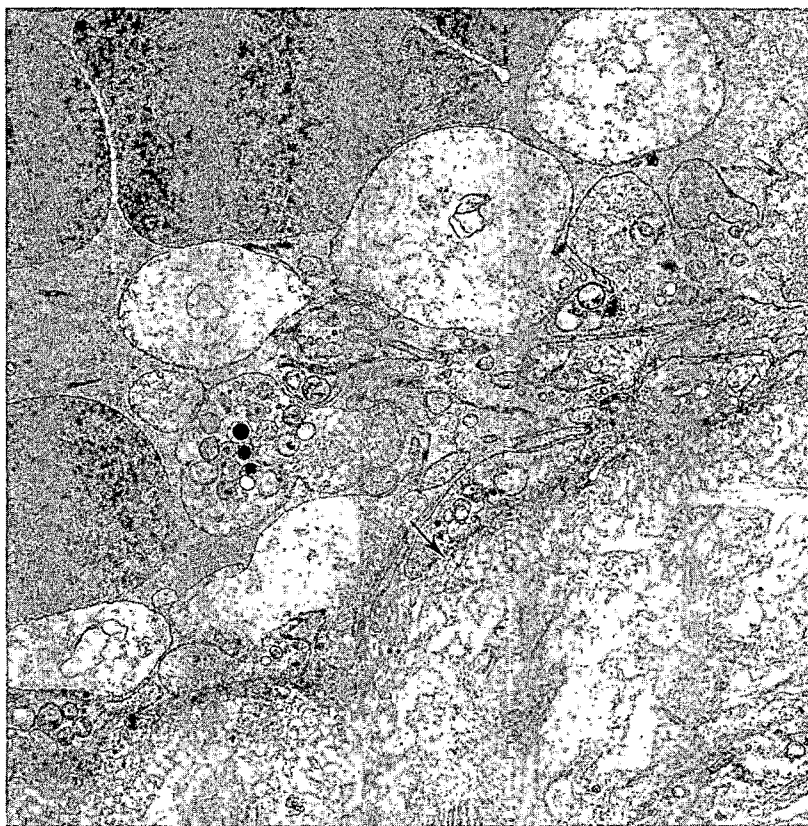


Figure 2 Twenty-four hours after endothelial removal, platelets adhere to the denuded surface and are seen in this electron micrograph to be closely apposed to the microfibrils (arrows) of the elastic fibers as well as to the basement membrane-like material. Fibrin is seen on the luminal side of the vessel between the platelets.

terolemia, nonhuman primates can be shown to have decreased platelet survival. After six days of chronic homocysteinemia baboons have as much as a 50% decrease in platelet survival. Autopsy revealed that as much as 10% of the endothelium of the thoracic and abdominal aorta was missing. In fact, Harker and his colleagues (32, 33) showed a linear correlation between the level of homocysteinemia and the amount of missing endo-



Figure 3 This low power electron micrograph represents a small segment of a three-month lesion in a pigtail monkey that was continuously hyperlipemic for the entire duration of the experiment. The deeper portion of the lesion contains numerous fat-filled smooth muscle cells. The extensive lipid deposits fill the cells and severely extend them. Many of these lipid deposits have been partially extracted and appear as central clear areas. In the more superficial portion of the lesion, numerous small deposits of extracellular lipid can be seen in the connective tissue matrix (*arrows*).

thelium in the aorta. There is also a correlation between the amount of missing endothelium and the decrease in platelet survival observed in these animals (Figure 4). If the animals were kept homocysteinemic for 90 days instead of 6 days, not only did they continue to manifest a decrease in platelet survival, but at the end of the 90 days there was a marked intimal smooth muscle proliferation at the sites of missing endothelium. The smooth muscle cells that had proliferated in these lesions were surrounded by large amounts of newly formed connective tissue. In addition, some of the lesions contained accumulations of intracellular lipid deep within the lesions, a characteristic commonly associated with an early fibrous plaque, even though the animals were not hypercholesterolemic (Figure 5).

Interestingly, the decrease in platelet survival was clearly correlated with atherosclerotic lesion formation, as demonstrated with pharmacologic agents inhibitory to platelet function. One of these, dipyridamole, returned platelet survival to normal levels in a group of chronically homocysteinemic baboons. Inhibition of platelet function with dipyridamole did not affect the amount of endothelium that was missing, yet it was specifically able to inhibit intimal smooth muscle proliferation. These studies were first to demonstrate that decreased platelet survival was associated with intimal

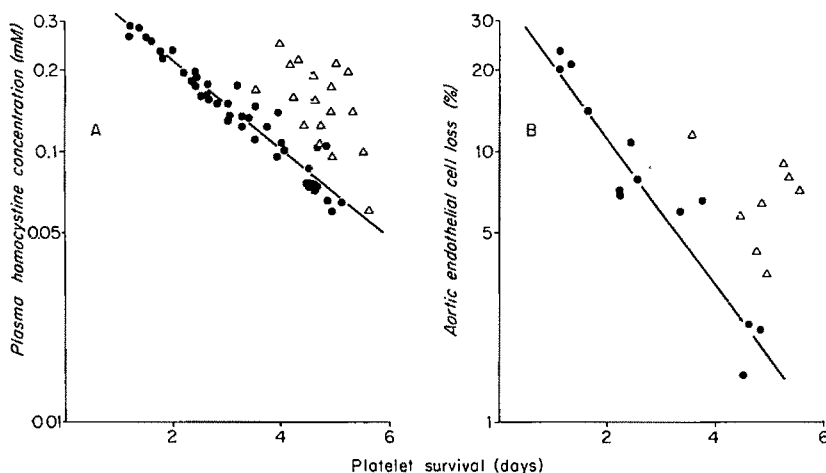


Figure 4 The relationships of platelet survival and plasma homocystine level and endothelial cell loss. (A) Platelet survival time correlated with the logarithm of the plasma homocystine concentration (dots) with $r = 0.965$, $P < 0.001$, and regression line represented by $y = 0.453(-0.368x)$. (B) Measurements of platelet survival time also correlated with the logarithm of endothelial cell loss (dots) with $r = 0.958$, $P < 0.001$, and a regression line represented by $y = 44.0(-0.647x)$. Dipyridamole treatment (triangles) produced significant prolongation of platelet survival ($P < 0.001$). Reproduced with permission (33).

smooth muscle proliferation and that, when platelet function was effectively inhibited, it was possible to prevent the development of the fibromusculo-proliferative lesions of atherosclerosis (33).

In similar fashion, Ross & Harker (21) observed that chronic hypercholesterolemia also leads to focal endothelial desquamation. They studied a series of pigtail monkeys made hypercholesterolemic by the addition of eggs and butter to their diet. These animals had serum cholesterol levels of approximately 250 mg per 100 ml and were missing as much as 5% of the endothelium in their thoracic and abdominal aortas. As before, this missing endothelium was associated with an approximate 40% reduction in platelet survival. In such chronically hypercholesterolemic monkeys, the sites of missing endothelium were associated with intimal lesion formation (Table 1).

The importance of the maintenance of endothelial integrity and the role of the platelet in the development of the lesions of atherosclerosis has been further substantiated by different approaches in at least three laboratories. Moore and his colleagues produced lesions in rabbits with an indwelling arterial catheter (30). The massive intimal smooth muscle proliferative lesion induced by such a catheter in normocholesterolemic animals is rich in lipid deposits. Moore and his colleagues (36) inhibited platelet function in these rabbits by inducing thrombocytopenia with an antiplatelet

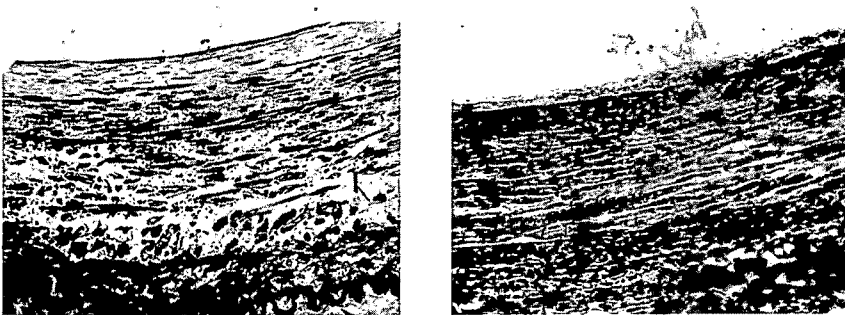


Figure 5 (A) A light micrograph demonstrating a characteristic lesion seen in the homocystinemic animals. The internal elastic lamina (*arrow*) is present at the bottom of this micrograph, and the lumen of the vessel at the top. It can be seen that the intima occupies essentially all of this micrograph, which demonstrates the marked increase in thickness, smooth muscle proliferation, and connective tissue formation that has occurred. Several lipid-laden cells (1) are present near the base of this lesion. A small portion of the media is present at the bottom of the micrograph. *(B)* A light micrograph of a segment of the iliac artery of a homocystinemic dipyrindamole-treated animal. It can be seen that the intima is very thin and that the media and adventitia are both visible compared with the markedly thickened intima demonstrated in the homocystinemic animals in *(A)*.

Table 1 Platelet kinetics in normal and hyperlipidemic monkeys

Diet	Animals (no.)	Plasma lipids		Endothelial cell loss (% of surface)	Platelet		
		Cholesterol (mg/dl)	Triglyceride (mg/dl)		Count (no./ml)	Survival (days)	Turnover of platelets (ml ⁻¹ day ⁻¹)
Normal	8	88 ± 5.3 ^a	30 ± 9.4	0	383,000 ± 62,000	8.0 ± 0.34	61,000 ± 11,000
Hyper- lipidemic	6	223 ± 22	28 ± 14	5.0 ± 1.2	396,000 ± 70,000	5.8 ± 0.54	86,000 ± 13,000
Probability		<0.01	>0.75	>0.001	>0.75	<0.01	<0.05

^aThe variation is ± 1 standard deviation.

serum. Such thrombocytopenic rabbits with indwelling catheters contained no smooth muscle proliferative response. Similarly, Friedman and his colleagues (37) were able to prevent the lesions induced by the intra-arterial balloon catheter in rabbits made thrombocytopenic by similar methods.

Of particular interest are the experiments of Fuster and his colleagues (38). They have been studying atherosclerosis in a group of swine that are homozygous for von Willebrand's disease as contrasted with a matched group of swine, both groups being fed hypercholesterolemia-inducing diets. Platelets in the von Willebrand swine are unable to undergo normal adherence and release upon exposure to collagen (or other substrates), presumably because of the absence of factor VIII in their plasma. The hypercholesterolemic swine with von Willebrand's disease demonstrated essentially no atherosclerotic lesions. In contrast, the matched hypercholesterolemic swine had extensive atherosclerotic lesions (38).

In each of these instances, the common denominators have been alterations or "injury" to the endothelium associated with decreased platelet survival (or increased platelet utilization) correlated with focal intimal smooth muscle proliferation. When platelet function is inhibited, the development of the proliferative smooth muscle lesions is then prevented. The question, of course, arises: How and why is platelet function associated with smooth muscle proliferation? The answer comes from a series of correlated *in vitro* studies of the factors that control the growth of a number of cells, including arterial smooth muscle cells, in cell culture.

THE PLATELET-DERIVED GROWTH FACTOR AND CELL PROLIFERATION

Most cells in culture require a factor present in whole blood serum to proliferate. In a series of studies, Ross and his colleagues (39-41) demonstrated that this growth is due to the activity of a mitogenic factor present in whole blood serum, which is derived from the process of platelet aggrega-

tion and release that normally occurs when serum is prepared. Thrombin is formed during the preparation of serum. Upon exposure to platelets, this thrombin leads to platelet aggregation and release. It is possible to make serum lacking in such platelet factors if the serum is made from cell-free plasma. Such cell-free, plasma-derived serum contains no mitogenic activity when tested upon a variety of cells including smooth muscle, fibroblasts, 3T3 cells, and glial cells. A striking and important exception to this observation is the fact that endothelial cells do not respond to the platelet factor and will grow logarithmically in the presence of an appropriate concentration of serum, regardless of whether the serum is derived from cell-free plasma or from whole blood (42, 43).

Thus, one of the principal mitogens present in whole blood serum and missing in cell-free, plasma-derived serum is a factor derived from the process of platelet aggregation and release. This factor can act upon smooth muscle cells but apparently not upon endothelium. The implications of these findings are important not only in relation to control of growth of cells in culture but in those *in vivo* proliferative lesions such as atherosclerosis where platelet function can be shown to be important as well.

The nature of this platelet-derived growth factor has been studied in several laboratories (44-47). It has a relatively low molecular weight (10,000-30,000), is a markedly cationic protein with a pI of approximately 9 to 10. The molecule is relatively heat stable (56°C for 30 minutes) and is extremely potent, since it is active upon cells in culture at levels of 100 ng/ml (or less) of culture medium (Figure 6). Platelets contain many substances; however, this is the only one that has thus far been identified as a mitogen. Of course, plasma also contains a number of hormones and substances that may be important in the growth of cells in culture. However, in the absence of platelet-derived growth factor, plasma will not support the proliferation of smooth muscle cells.

The growth of endothelial cells in culture appears to be under a different set of controls, since as indicated earlier endothelial cells will proliferate in the absence of the platelet-derived growth factor as long as an appropriate concentration of plasma is provided in the culture medium. Once the studies of the isolation, purification, and characterization of the platelet-derived growth factor have been completed, it should be possible to analyze its mode of action at the cellular and molecular level.

These studies in cell culture provide clear evidence that platelets can deliver a mitogenic substance *in vivo* to cells that can respond to it by proliferating and thus participate in the process of atherogenesis. The fact that the platelet mitogen is extremely potent suggests that only small quantities of it need be present to induce a relatively massive proliferative response on the part of the exposed cells.

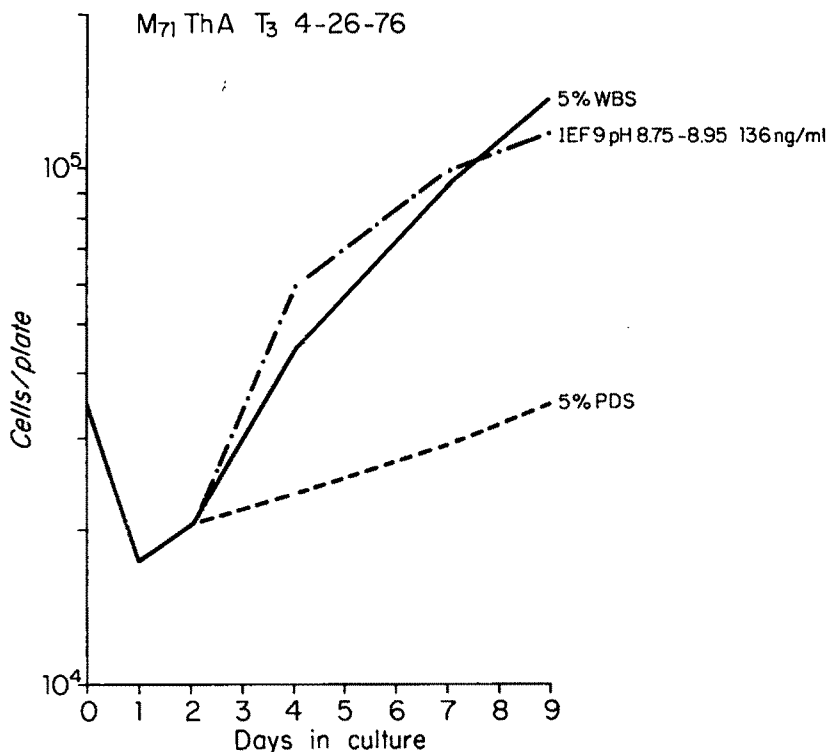


Figure 6 Response of arterial smooth muscle cells (SMC) to platelet-derived growth factor: 3×10^4 SMC were added to a large series of 35-mm petri dishes in 0.5% whole blood serum (WBS). After two days the dishes were separated into three groups. The cells fed 5% cell-free, plasma-derived serum (PDS) remained essentially quiescent. The cells fed either 5% WBS or 5% PDS plus 136 ng/ml of partially purified platelet factor (IEF), purified by isoelectric focusing from previously chromatographed fractions, grew logarithmically for a period of one week. These experiments demonstrate that even the partially purified platelet factor is a very potent mitogen.

CONCLUSION

The opportunity to study the biology of endothelium and smooth muscle cells in vivo and in cell culture has permitted us to test a hypothesis and provide further evidence for the interaction between platelets and smooth muscle cells in the induction of the intimal proliferative response that can lead to atherosclerosis. The maintenance of endothelial integrity appears to be a key factor in the prevention of atherogenesis. Injury to the endothelium can lead to a sequence of events that permits platelet- and plasma-derived molecules to gain entry into the artery wall and to stimulate intimal smooth

muscle proliferation. The discovery of an extremely potent molecule, with hormone-like qualities, that is present in the platelets and that can, upon exposure to smooth muscle cells, induce a massive proliferative response in vitro represents an important link between the studies of experimentally induced atherosclerosis and experiments studying control of cell proliferation in culture.

Based upon these studies, approaches to prevention and therapy of atherosclerosis should include development of means to protect the endothelial cells against injury, and of methods to inhibit platelet function that do not compromise the role of the platelet in blood coagulation. Future research should explain the means by which each of the risk factors commonly associated with atherosclerosis influences the course of this disease process.

ACKNOWLEDGEMENT

The author would like to acknowledge the collaboration and help of John Glomset, Laurence A. Harker, Arthur A. Vogel, Peter Davies, Beverly Kariya, Elaine Raines, and Mary-Jane Rivest in these studies. This work was supported in part by grants AM 13970, HL 18645, and AG 00299 from the USPHS and grant RP 00266 from the Regional Primate Center.

Literature Cited

1. *Arteriosclerosis: A Report by the National Heart and Lung Institute Task Force on Arteriosclerosis*. 1971. DHEW Publ. No. (NIH) 72-219, Vol. 2. Washington, DC: GPO
2. McNamara, J. J., Malot, M. A., Stremple, J. F., Cutting, R. T. 1971. Coronary artery disease in combat casualties in Vietnam. *J. Am. Med. Assoc.* 216: 1185-87
3. Enos, W. F., Holmes, R. H., Beyer, J. 1953. Coronary disease among United States soldiers killed in action in Korea. *J. Am. Med. Assoc.* 152:1090-93
4. Ross, R., Glomset, J. A. 1976. The pathogenesis of atherosclerosis. *N. Engl. J. Med.* 295:369-77, 420-25
5. Geer, J. C., McGill, H. C. Jr., Strong, J. P. 1968. The fine structure of human atherosclerotic lesions. *Am. J. Pathol.* 33:263-87
6. Geer, J. C., Haust, M. D. 1972. Smooth muscle cells in atherosclerosis. *Monogr. Atheroscler.* 2:1-88
7. Wissler, R. W. 1974. In *The Myocardium: Failure and Infarction*, ed. E. Braunwald, pp. 155-66. New York: Hospital Practice
8. Bierman, E. L., Ross, R. 1977. *Atheroscler. Rev.* 2:79-111
9. Kannel, W. B., LeBauer, E. J., Dawber, T. R., McNamara, P. M. 1967. Relation of body weight to development of coronary heart disease: The Framingham study. *Circulation* 35:734-44
10. Dawber, T. R. 1975. In *Current Concepts*, pp. 1-6. Kalamazoo, Mich: Upjohn
11. Johnson, B. C., Epstein, F. H., Kjelsberg, M. O. 1965. Distributions and familial studies of blood pressure and serum cholesterol levels in a total community—Tecumseh, Michigan. *J. Chronic Dis.* 18:147-60
12. Ostrander, L. D. Jr., Neff, B. J., Block, W. D., Francis, T. Jr., Epstein, F. H. 1967. Hyperglycemia and hypertriglyceridemia among persons with coronary heart disease. *Ann. Intern. Med.* 67:34
13. Goldstein, J. L., Hazzard, W. R., Schrott, H. G., Bierman, E. L., Motulsky, A. G. 1973. Hyperlipidemia in coronary heart disease. I. Lipid levels in 500 survivors of myocardial infarction. *J. Clin. Invest.* 52:1533-43

14. French, J. E. 1966. Atherosclerosis in relation to the structure and function of the arterial intima, with special reference to the endothelium. *Int. Rev. Exp. Pathol.* 5:253-353
15. Wright, H. P. 1972. Mitosis patterns in aortic endothelium. *Atherosclerosis* 15:93-100
16. Schwartz, S. M., Benditt, E. P. 1973. Cell replication in the aortic endothelium: A new method for study of the problem. *Lab. Invest.* 28:699-707
17. McGill, H. C. Jr., Geer, J. C., Strong, J. P. 1963. In *Atherosclerosis and Its Origin*, ed. M. Sandler, G. H. Bourne, pp. 39-65. New York: Academic
18. Haust, M. D. 1968. Electron microscopic and immunohistochemical studies of fatty streaks in human aorta. *Prog. Biochem. Pharmacol.* 4:429-37
19. Fry, D. L. 1973. Responses of the arterial wall to certain physical factors. *Ciba Found. Symp.* 12:93-125
20. Ross, R., Glomset, J. 1973. Atherosclerosis and the arterial smooth muscle cell. *Science* 180:1332-39
21. Ross, R., Harker, L. 1976. Hyperlipidemia and atherosclerosis. *Science* 193:1094-1100
22. Virchow, R. 1856. *Phlogose und Thrombose im Gefässsystem, Gesammelte Abhandlungen zur Wissenschaftlichen Medizin*, p. 458. Frankfurt-am-Main: Meidinger
23. Ross, R. 1971. The smooth muscle cell. II. Growth of smooth muscle in culture and formation of elastic fibers. *J. Cell Biol.* 50:172-86
24. Ross, R. 1973. In *Proc. Sigrid Juselius Found. Symp. in Turku, Finland*, ed. E. Kulonen, pp. 627-36. London: Academic
25. Narayanan, S., Sandberg, L. B., Ross, R., Layman, D. 1976. The smooth muscle cell. III. Elastin synthesis in arterial smooth muscle cell culture. *J. Cell Biol.* 68:411-19
26. Wight, T. N., Ross, R. 1975. Proteoglycans in primate arteries. II. Synthesis and secretion of glycosaminoglycans by arterial smooth muscle cells in culture. *J. Cell Biol.* 67:675-86
27. Burke, J. M., Balian, G., Ross, R., Bornstein, P. 1977. Synthesis of types I and III collagen and procollagen by monkey aortic smooth muscle cells in vitro. *Biochemistry* 16:3243-49
28. Stermerman, M. B., Ross, R. 1972. Experimental arteriosclerosis. I. Fibrous plaque formation in primates, an electron microscope study. *J. Exp. Med.* 136:769-89
29. Baumgartner, H. R., Studer, A. 1966. Folgen des Gefäßkatheterismus am normo- und hyperscholesterinaemischen Kaninchen. *Pathol. Microbiol.* 29:393-405
30. Moore, S. 1973. Thromboatherosclerosis in normolipemic rabbits: A result of continued endothelial damage. *Lab. Invest.* 29:478-87
31. Friedman, R. J., Moore, S., Singal, D. P. 1975. Repeated endothelial injury and induction of atherosclerosis in normolipemic rabbits by human serum. *Lab. Invest.* 32:404-15
32. Harker, L. A., Slichter, S. J., Scott, C. R., Ross, R. 1974. Homocysteinemia—vascular injury and arterial thrombosis. *N. Eng. J. Med.* 291:537-43
33. Harker, L., Ross, R., Slichter, S., Scott, C. 1976. Homocystine-induced arteriosclerosis: The role of endothelial cell injury and platelet response in its genesis. *J. Clin. Invest.* 58:731-41
34. Minick, C. R., Murphy, G. E., Campbell, W. G. Jr. 1966. Experimental induction of athero-arteriosclerosis by the synergy of allergic injury to arteries and lipid-rich diet. I. Effect of repeated injections of horse serum in rabbits fed a dietary cholesterol supplement. *J. Exp. Med.* 124:635-52
35. Minick, C. R., Murphy, G. E. 1973. Experimental induction of atheroarteriosclerosis by the synergy of allergic injury to arteries and lipid-rich diet. II. Effect of repeated injections of horse serum in rabbits fed a lipid-rich, cholesterol-poor diet. *Am. J. Pathol.* 73:265-300
36. Moore, S., Friedman, R. J., Singal, D. P., Gauldie, J., Blajchman, M. 1976. Inhibition of injury induced thromboatherosclerotic lesions by antiplatelet serum in rabbits. *Thromb. Diath. Haemorrh.* 35:70-81
37. Friedman, R. J., Stermerman, M. B., Wenz, B., Moore, S., Gauldie, J., Gent, M., Tiell, L., Spaet, T. H. 1977. The affect of thrombocytopenia on experimental arteriosclerotic lesion formation in rabbits. *J. Clin. Invest.* 60:1191-1201
38. Fuster, V., Bowie, E. J. W., Lewis, J. C., Fass, D. N., Owen, C. A. Jr., Brown, A. L. 1978. Resistance to arteriosclerosis in pigs with von Willebrand's disease. *J. Clin. Invest.* 61:722-30
39. Ross, R., Glomset, J., Kariya, B., Harker, L. 1974. A platelet-dependent serum factor that stimulates the proliferation of arterial smooth muscle cells in vitro. *Proc. Natl. Acad. Sci. USA* 71:1207-10

40. Rutherford, B. R., Ross, R. 1976. Platelet factors stimulate fibroblasts and smooth muscle cells quiescent in plasma-serum to proliferate. *J. Cell Biol.* 69:196-204
41. Ross, R., Vogel, A. 1978. The platelet-derived growth factor. *Cell* 14:203-10
42. Wall, R. T., Harker, L. A., Quadracci, L. J., Striker, G. E. 1978. Factors influencing endothelial cell proliferation in vitro. *J. Cell Physiol.* In press
43. Davies, P. F., Ross, R. 1978. Mediation of pinocytosis in cultured arterial smooth muscle and endothelial cells by platelet-derived growth factor. *J. Cell Biol.* In press
44. Antoniades, H. N., Scher, C. D. 1977. Radioimmunoassay of a human serum growth factor for Balb/c-3T3 cells: Derivation from platelets. *Proc. Natl. Acad. Sci. USA* 74:1973-77
45. Antoniades, H. N., Scher, C. D. 1978. Growth factors derived from human serum, human platelets and human pituitary: Properties and immunologic crossreactivity. *J. Natl. Cancer Inst.* In press
46. Heldin, C.-H., Wasteson, A., Westermark, B. 1977. Partial purification and characterization of platelet factors stimulating the multiplication of normal human glial cells. *Exp. Cell Res.* 109:429-37
47. Kohler, N., Lipton, A. 1974. Platelets as a source of fibroblast growth-promoting activity. *Exp. Cell Res.* 87:297-301
48. Ritchie, J. L., Harker, L. A. 1977. Platelet and fibrinogen survival in coronary atherosclerosis: Response to medical and surgical therapy. *Am. J. Cardiol.* 39:595-98

HAIRY CELL LEUKEMIA

♦7303

Raul C. Braylan, M.D.

Department of Pathology, University of Florida College of Medicine,
Gainesville, Florida 32610

Jerome S. Burke, M.D.

Department of Pathology, Stanford University Medical Center,
Stanford, California 94305

INTRODUCTION

Hairy cell leukemia, or leukemic reticuloendotheliosis, is a chronic leukemia that has received increased attention in recent years. Much of the current vogue in the recognition of this disease may be due to the employment of the term "hairy cell," first used by Schrek & Donnelly (1) to describe the villous cytoplasmic projections of the leukemic cells as viewed under phase microscopy. Because of the curious nature of the hairy cell, there has been considerable interest in attempting to define the cytogenesis, natural history, and factors necessary for accurate diagnosis and effective therapy of this disease.

CLINICAL AND LABORATORY FEATURES

Although termed "leukemia," paradoxically the majority of patients with hairy cell leukemia are aleukemic and present with pancytopenia and clinical hypersplenism (2-6). The disease tends to affect males aged 50-60, and is relatively less frequent in females. Initial symptoms such as fatigue, lassitude, infections, bruising, or abdominal fullness are related to the pancytopenia or enlarged spleen. Many patients are, in fact, asymptomatic and the disease is only discovered on routine hematologic survey or physical examination. The most striking physical finding is an enlarged, frequently massive spleen (2-6). Lymphadenopathy is distinctly uncommon; hepatomegaly generally is moderate.

Blood studies usually reveal a normochromic normocytic anemia, thrombocytopenia, and leukopenia averaging $2-3 \times 10^3$ white blood cells/mm³ with a relative lymphocytosis (2-7). Overt leukemia with white blood cell counts greater than 10×10^3 /mm³ is rare at presentation (2-4, 7). Hairy cells frequently are difficult to recognize in peripheral blood smears (Figure 1) and are commonly misdiagnosed as "atypical" lymphocytes (7). In some cases no hairy cells can be identified in the blood. If hairy cells are observed or suspected, the tartrate-resistant acid phosphatase reaction (TRAP) is a useful cytochemical test to confirm the diagnosis (8). It should be realized that the TRAP, which marks for isoenzyme 5, is not completely specific and the intensity of positivity may be extremely variable (9). In rare cases the reaction may be negative (2, 3). Bone marrow aspirates generally result in dry taps (2, 3, 7, 9) but in our experience marrow biopsy is one of the most reliable diagnostic procedures.

The results of other clinical and laboratory studies have not been consistently abnormal. Except for one reported case of monoclonal gammopathy (10), immunoelectrophoretic studies are essentially unremarkable (2, 4). Serum muramidase levels are generally normal or slightly depressed (2, 3), and chromosomal analysis has not been fruitful except for one recent study reporting loss of Y chromosomes in 2 of 20 patients studied (11). Investigation of platelet aggregation revealed some functional abnormalities (12, 13),

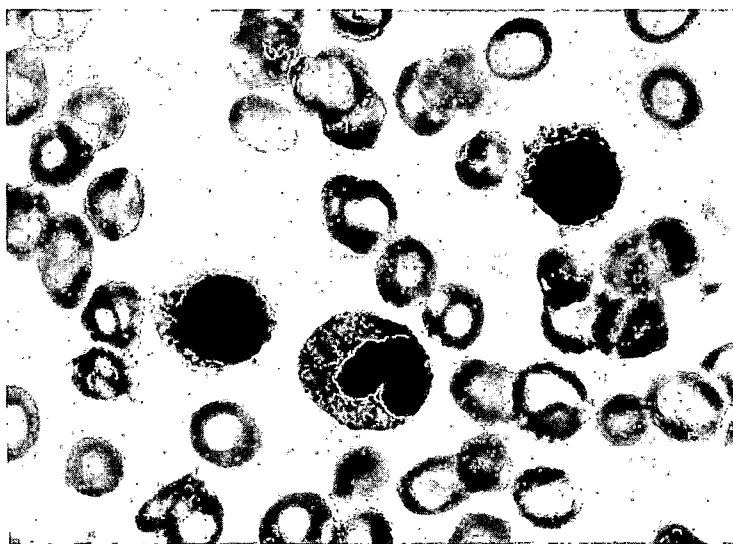


Figure 1 Hairy cells in peripheral blood. Note the size in comparison to the band granulocyte ($\times 1200$).

but intrinsic red cell defects have not been described. Red cell survival studies using ^{51}Cr show increased sequestration of erythrocytes in the spleen (2, 3) and quantitative scanning reveals that the splenic red cell volume is generally increased, particularly in comparison to myelo- and lymphoproliferative disorders with similar splenomegaly (14).

PATHOLOGY

Bone marrow and spleen are usually heavily involved by the leukemic cells and are the most common tissues available to the pathologist for diagnosis. Liver and lymph nodes may also be of diagnostic value. The marrow may be partially or completely replaced (7). In tissue sections the hairy cells appear homogeneous and bland without mitoses or large nucleoli, and are characterized by a distinct water-clear zone of cytoplasm around the cell nuclei that imparts a halo appearance (Figure 2). This is in contrast to the lack of visible cytoplasm and close apposition of nuclei in chronic lymphocytic leukemia, most often confused with hairy cells in biopsies (7, 15). Marrow reticulin is often increased, which may account for the frequency

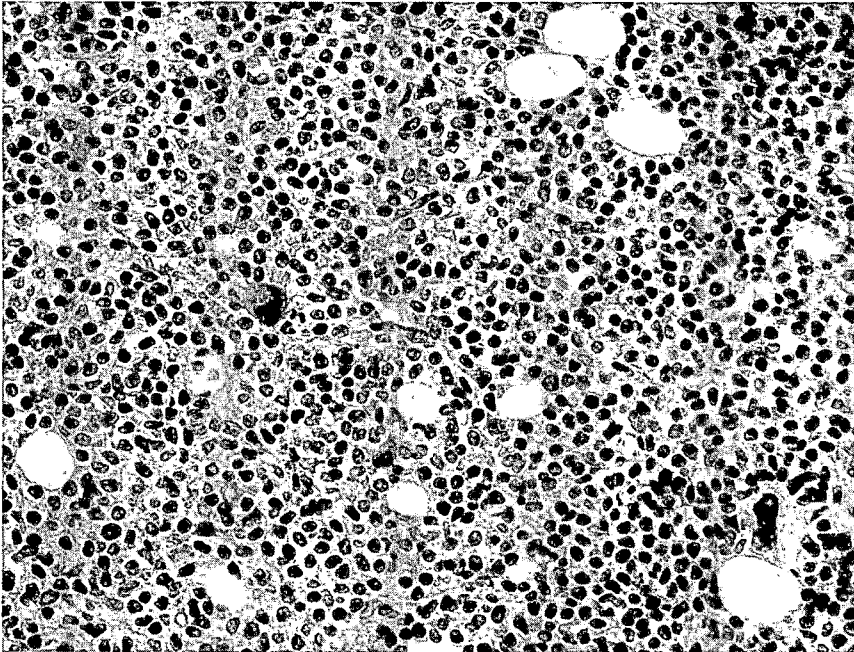


Figure 2 Bone marrow infiltrated by hairy cell leukemia with the characteristic clear zones around the nuclei. Residual megakaryocytes are present (X 300).

of poor yields on aspiration (6, 7, 15). Imprint smears are a useful adjunct to the biopsy, particularly when coupled with staining for TRAP (16). As in the bone marrow, the splenic histology of hairy cell leukemia is distinctive (2, 6, 17); the splenic red pulp, cords, and sinuses are diffusely infiltrated by a uniform population of monotonous-appearing hairy cells resulting in reduction in the size of the white pulp (2). Blood-filled spaces of variable size in the red pulp (blood "lakes") are often seen (17), which perhaps provides a morphologic explanation for the increased erythrocyte pooling (14). In addition to hairy cells and erythrocytes, histochemical studies of the spleen reveal increased numbers of macrophages in the pulp cords (18), which supports the ultrastructural observation of leukoerythrophagocytosis by splenic macrophages (19). The increased phagocytic activity of apparently normal macrophages may account, at least in part, for the hypersplenic syndrome in hairy cell leukemia. Transmission electron microscopic studies may also be of value in the diagnosis of this disease since in many instances the leukemic cells contain a characteristic cytoplasmic ribosomal-lamellar complex (20, 21). Although these complexes are frequently observed in hairy cell leukemia, they are not pathognomonic and have been described in other disorders (22).

THERAPY AND PROGNOSIS

Infection secondary to pancytopenia is the primary complication in hairy cell leukemia (4, 5, 23). The majority of studies indicate that splenectomy is the treatment of choice in hairy cell leukemia (2, 3, 5). However, the indication for splenectomy will depend upon the degree of pancytopenia and splenomegaly. For most patients splenectomy usually results in restoration of normal hemtologic parameters and/or sufficient clinical improvement to ameliorate the symptomatology. One series, however, reported either failure or only a partial reversion of the pancytopenia in some 40% of patients (3); these patients had a median survival approximately five times less than patients who responded completely to splenic removal. Nonetheless, the overall survival at five years in hairy cell leukemia is approximately 50% (2) and survivals over ten years have been recorded (2, 5). All patients with hairy cell leukemia, regardless of response to splenectomy, have persistent bone marrow involvement (7). It is possible that patients who do not respond to splenectomy have more severe marrow involvement, hence the impairment of marrow granulocyte reserve and leukocyte production (24). This is significant since the majority of patients with hairy cell leukemia who fail after splenectomy succumb to infectious complications (3, 5). In particular there has been recent indication of a curiously increased incidence of mycobacterial infections (3, 25), which may be related to the monocytopenia reported in this disorder (26). The

danger of infection adds a cautionary note to the use of chemotherapeutic agents in hairy cell leukemia. With few exceptions (3, 27), cytotoxic agents or steroids have not been beneficial either as initial therapy or following splenectomy. On the contrary, aggressive chemotherapy, generally employed because of a misdiagnosis of acute leukemia or lymphoma, often leads to dire consequences in the patient with hairy cell leukemia (2, 5). The failure of chemotherapy in controlling this disease may be related to the fact that hairy cells have a low proliferative capacity as has been shown by thymidine incorporation and flow cytometry (18). At present, no effective treatment is known for patients not responding to splenectomy. However, recent attempts to remove circulating hairy cells by leukapheresis in leukemic patients have met with some success (28, 29). Controlled prospective studies will be necessary to assess therapeutic alternatives for patients with progressive or late stage disease.

THE NATURE OF THE HAIRY CELL

The origin of the neoplastic cells of hairy cell leukemia is still highly controversial. Multiple cytochemical, structural, and functional techniques that define normal hematopoietic elements have been applied to hairy cells to characterize their nature, but no consistent results have been obtained (12, 18, 20, 30-39). Most studies suggested either a monocytic origin (35-37) or a B lymphocytic derivation (31, 32, 34, 38), and a few investigators consider that hairy cells may be B lymphocytes with phagocytic properties (12, 33, 39), or hybrid cells (30), or unique elements that are not detected in normal conditions (18). Hairy cells do not have high phagocytic capacity, although in vitro they have a limited ability to ingest particles or microorganisms (12, 18, 20, 33, 36, 39). Most observers found receptors for the Fc fragment of IgG (18, 30, 33, 35-37) on hairy cells but the detection of complement receptors varied widely (18). Structurally, hairy cells display ample surface ridges and ruffles (Figure 3) (18) that account for their "hairy" appearance. This complex surface may also contribute to their large electronic size as detected by Coulter analysis (18). The peculiar ruffled surface of hairy cells as observed by scanning microscopy may be seen in normal monocytes but it is not a feature of freshly isolated normal or neoplastic lymphocytes (18). In most reported instances, hairy cells demonstrated surface-associated immunoglobulins (SIg) (12, 18, 30-34, 36), which suggests a B cell lineage. Some observations indicate that at least in part the SIg may have been extrinsically bound to the cells and not an actual cell product (18, 33, 38). However, several studies demonstrated restriction of the SIg to a single light-chain determinant (12, 18, 30), and a few observations indicated the presence of cytoplasmic immunoglobulin (30, 34) and even immunoglobulin synthesis in hairy cells (12, 32, 34, 39), which would

2197 P 24, 840

conclusively demonstrate their B lymphocytic origin. Moreover, monoclonal serum immunoglobulin elevation in hairy cell leukemia has been reported (10). It must be emphasized, however, that these results have not been universally confirmed. The presence or synthesis of immunoglobulins in hairy cell populations has not always been observed (18, 40) and the finding of a serum paraprotein in hairy cell leukemia is an exceedingly rare event (2, 4). These differences in results may be due to a lack of standardized methods to assay the different functional characteristics of hairy cells. It is also reasonable to assume that these cells, although morphologically and cytochemically homogeneous, vary in their expression of such characteristics. However, an important source of discrepancy may result from the lack of strict diagnostic criteria to define hairy cell leukemia. There is evidence that some of the clinical and hematological criteria that are used to diagnose hairy cell leukemia may be seen in disorders that differ histologically and biologically from conventional hairy cell leukemia. That is, typical B cell lymphoproliferative disorders with a monoclonal serum paraprotein and with cells containing cytoplasmic immunoglobulins may present with splenomegaly, pancytopenia, and TRAP-positive hairy cells (40, 41). Obviously, these findings raise questions about the validity of some functional studies in hairy cell leukemia and stress the necessity of reliable and uniform criteria for defining and diagnosing this disease.

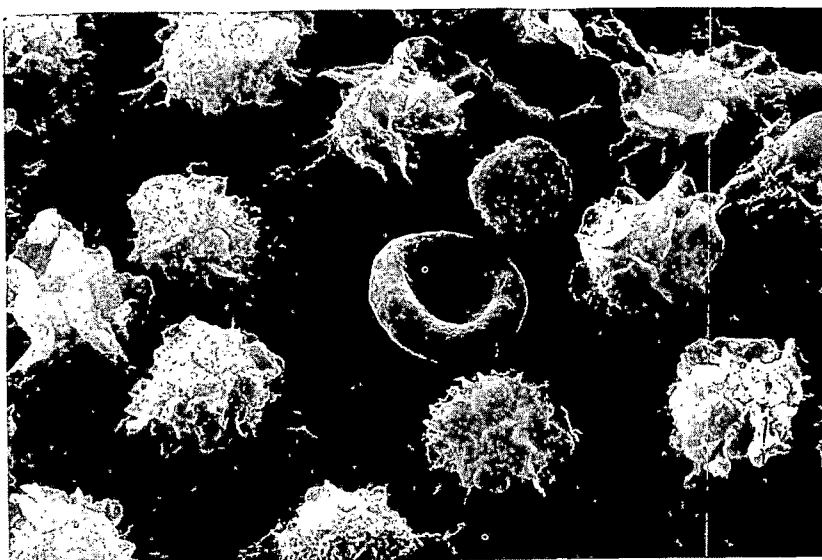


Figure 3 Hairy cells as observed by scanning electron microscopy. Note the ample surface ruffles. An erythrocyte is seen in the middle of the field (X 3700).

Literature Cited

- Schrek, R., Donnelly, W. J. 1966. "Hairy" cells in blood in lymphoreticular neoplastic disease and "flagellated" cells of normal lymph nodes. *Blood* 27:199-211
- Burke, J. S., Byrne, G. E. Jr., Rappaport, H. 1974. Hairy cell leukemia (Leukemic reticuloendotheliosis). I. A clinical pathologic study of 21 patients. *Cancer* 33:1399-1410
- Catovsky, D. 1977. Hairy-cell leukaemia and prolymphocytic leukaemia. *Clin. Haematol.* 6:245-68
- Flandrin, G., Daniel, M. T., Fourcade, M., Chelloul, N. 1973. Leucemie a "tricholeucocyte" (hairy cell leukemia). Etude clinique et cytologique de 55 observations. *Nouv. Rev. Fr. Hematol.* 13:609-40
- Katayama, I., Finkel, H. E. 1974. Leukemic reticuloendotheliosis. A clinicopathologic study with review of the literature. *Am. J. Med.* 57:115-26
- Naeim, F., Smith, G. S. 1974. Leukemic reticuloendotheliosis. *Cancer* 34:1813-21
- Burke, J. S. 1978. The value of the bone marrow biopsy in the diagnosis of hairy cell leukemia. *Am. J. Clin. Pathol.* In press
- Yam, L. T., Li, C. Y., Lam, K. W. 1971. Tartrate-resistant acid phosphatase isoenzyme in the reticulum cells of leukemic reticuloendotheliosis. *N. Engl. J. Med.* 284:357-60
- Katayama, I., Yang, J. P. S. 1977. Reassessment of a cytochemical test for differential diagnosis of leukemic reticuloendotheliosis. *Am. J. Clin. Pathol.* 68:268-72
- Golde, D. W., Saxon, A., Stevens, R. H. 1977. Macroglobulinemia and hairy cell leukemia. *N. Engl. J. Med.* 296:92-93
- Golomb, H. M., Lindgren, V., Rowley, J. D. 1978. Chromosome abnormalities in patients with hairy cell leukemia. *Cancer* 41:1374-80
- Golomb, H. M., Vardiman, J., Sweet, D. L. Jr., Simon, D., Variakojis, D. 1978. Hairy cell leukaemia: Evidence for the existence of a spectrum of functional characteristics. *Br. J. Haematol.* 38:161-70
- Levin, P. H., Katayama, I. 1975. The platelet in leukemic reticuloendotheliosis. Functional and morphological evidence of a qualitative disorder. *Cancer* 36:1353-58
- Lewis, S. M., Catovsky, D., Hows, J. M., Ardalán, B. 1977. Splenic red cell pooling in hairy cell leukaemia. *Br. J. Haematol.* 35:351-57
- Vykoupil, K. F., Thiele, J., Georgii, A. 1976. Hairy cell leukemia. Bone marrow findings in 24 patients. *Virchows Arch. A* 370:273-89
- Krause, J. R., Srodes, C., Lee, R. E. 1977. Use of the bone marrow imprint in the diagnosis of leukemic reticuloendotheliosis ("hairy cell leukemia"). *Am. J. Clin. Pathol.* 68:368-71
- Nanba, K., Soban, E. J., Bowling, M. C., Berard, C. W. 1977. Splenic pseudosinusoids and hepatic angiomatous lesions. Distinctive features of hairy cell leukemia. *Am. J. Clin. Pathol.* 67:415-26
- Braylan, R. C., Jaffe, E. S., Triche, T. J., Nanba, K., Fowlkes, B. J., Metzger, H., Frank, M. M., Dolan, M. S., Yee, C. L., Green, I., Berard, C. W. 1978. Structural and functional properties of the "hairy" cells of leukemic reticuloendotheliosis. *Cancer* 41:210-27
- Burke, J. S., Mackay, B., Rappaport, H. 1976. Hairy cell leukemia (Leukemic reticuloendotheliosis). II. Ultrastructure of the spleen. *Cancer* 37:2267-74
- Daniel, M. T., Flandrin, G. 1974. Fine structure of abnormal cells in hairy cell (tricholeukocytic) leukemia, with special reference to their in vitro phagocytic capacity. *Lab. Invest.* 30:1-8
- Katayama, I., Schneider, G. B. 1977. Further ultrastructural characteristics of hairy cells of leukemic reticuloendotheliosis. *Am. J. Pathol.* 86:163-82
- Brunning, R. D., Parkin, J. 1975. Ribosome-lamella complexes in neoplastic hematopoietic cells. *Am. J. Pathol.* 79:565-78
- Bouza, E., Burgaleta, C., Golde, D. W. 1978. Infections in hairy-cell leukemia. *Blood* 51:851-59
- Yam, L. T., Chaudhry, A. A., Janckila, A. J. 1977. Impaired marrow granulocyte reserve and leukocyte mobilization in leukemic reticuloendotheliosis. *Ann. Intern. Med.* 87:444-46
- Marie, J. P., Degos, L., Flandrin, G. 1977. Hairy-cell leukemia and tuberculosis. (Letter.) *N. Engl. J. Med.* 297:1354
- Seshadri, R. S., Brown, E. J., Zipursky, A. 1976. Leukemic reticuloendotheliosis. A failure of monocyte production. *N. Engl. J. Med.* 295:181-84
- Davies, T. E., Waterbury, L., Abeloff, M., Burke, P. J. 1976. Leukemic reticuloendotheliosis: Report of a case with prolonged remission following in-

- tensive chemotherapy. *Arch. Intern. Med.* 136:620-22
28. Keitt, A. S., Weiner, R. S. 1978. Therapeutic leukapheresis in hairy cell leukemia. *Clin. Res.* 26:52A (Abstr.)
 29. Moore, J. O., Fay, J., Logue, G. L. 1977. Hairy cell leukemia: Proliferative potential and treatment with intensive leukopheresis. *Blood* 50:156 (Abstr.)
 30. Burns, G. F., Nash, A. A., Worman, C. P., Barker, C. R., Hayhoe, F. G. J. 1977. A human leukaemic cell expressing hybrid membrane phenotypes. *Nature* 268:243-45
 31. Catovsky, D., Pettit, J. E., Galetto, J., Okos, A., Galton, D. A. G. 1974. The B-lymphocyte nature of the hairy cell of leukaemic reticuloendotheliosis. *Br. J. Haematol.* 26:29-37
 32. Debusscher, L., Bernheim, J. L., Colard-Ronge, E., Govaerts, A., Hooghe, R., Lejeune, F. J., Zeicher, M., Stryckmans, P. A. 1975. Hairy cell leukemia: Functional, immunologic, kinetic, and ultrastructural characterization. *Blood* 46:495-507
 33. Fu, S. M., Winchester, R. J., Rai, K. R., Kunkel, H. G. 1974. Hairy cell leukemia: Proliferation of a cell with phagocytic and B-lymphocyte properties. *Scand. J. Immunol.* 3:847-51
 34. Golde, D. W., Stevens, R. H., Quan, S. G., Saxon, A. 1977. Immunoglobulin synthesis in hairy cell leukaemia. *Br. J. Haematol.* 35:359-65
 35. Jaffe, E. S., Shevach, E. M., Frank, M. M., Green, I. 1974. Leukemic reticuloendotheliosis: Presence of a receptor for cytophilic antibody. *Am. J. Med.* 57:108-14
 36. King, G. W., Hurtubise, P. E., Sagone, A. L. Jr., LoBuglio, A. F., Metz, E. N. 1975. Leukemic reticuloendotheliosis. A study of the origin of the malignant cell. *Am. J. Med.* 59:411-16
 37. Scheinberg, M., Brenner, A. I., Sullivan, A. L., Cathcart, E. S., Katayama, I. 1976. The heterogeneity of leukemic reticuloendotheliosis, "hairy cell leukemia." Evidence for its monocytic origin. *Cancer* 37:1302-7
 38. Stein, H., Kaiserling, E. 1974. Surface immunoglobulins and lymphocyte-specific surface antigens on leukaemic reticuloendotheliosis cells. *Clin. Exp. Immunol.* 18:63-71
 39. Utsinger, P. D., Yount, W. J., Fuller, C. R., Logue, M. J., Orringer, E. P. 1977. Hairy cell leukemia: B-lymphocyte and phagocytic properties. *Blood* 49:19-27
 40. Neiman, R. S., Sullivan, A., Jaffe, R. 1978. Malignant lymphoma simulating leukemic reticuloendotheliosis. *Cancer*. In press
 41. Katayama, I. 1977. Hairy cell leukemia. (Letter.) *N. Engl. J. Med.* 296:881

BIOLOGICAL AND ENZYMATIC EVENTS IN CHEMICAL CARCINOGENESIS

❖7304

Henry C. Pitot, M.D., Ph.D.

McArdle Laboratory for Cancer Research, Departments of Oncology and
Pathology, Medical School, University of Wisconsin, Madison, Wisconsin 53706

INTRODUCTION

That chemicals and ionizing radiation induce cancer in mammals was first observed in the human. Percival Pott, an eminent English surgeon of the 18th century, published the first clear demonstration of an occupational cancer resulting from chronic exposure to a specific environmental agent, soot. A century later, soon after the discovery of x rays by Roentgen, "radiation cancers" in humans were reported in scientists who utilized radioactive sources. More than a decade later, in 1908, Clunet demonstrated the experimental production of neoplasia by irradiation (1).

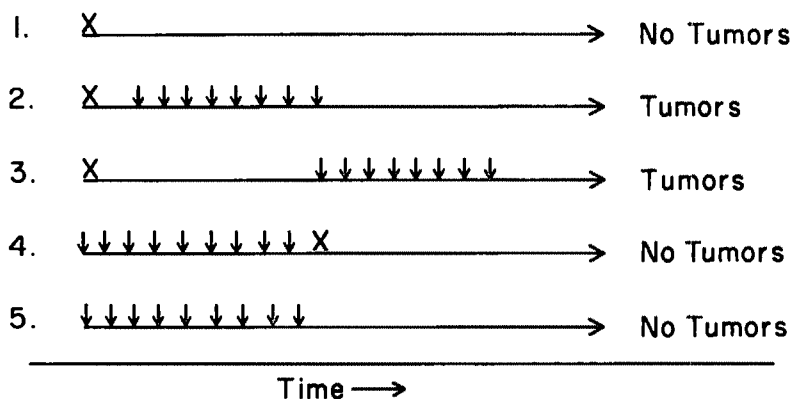
Although 200 years have passed since Dr. Pott's report, it has only been within the last three decades that our knowledge of the natural history and biochemistry of chemical carcinogenesis has progressed beyond that of descriptive pathology. The critical experiments allowing for our expanded knowledge of the biology and chemistry of carcinogenesis can be traced to the investigations of Berenblum & Shubik (2, 3) on the natural history of skin carcinogenesis in mice, and to the brilliant experiments of Drs. James and Elizabeth Miller and their associates (4, 5) on the metabolism of chemical carcinogens. This brief review acquaints the reader with the knowledge that stems from the critical experiments cited above of the biological and enzymatic events in chemical carcinogenesis.

THE NATURAL HISTORY OF CHEMICALLY INDUCED NEOPLASIA

The systematic study of the natural history of the neoplastic process *in vivo* dates principally from the pioneering studies of Berenblum & Shubik (2, 3), who demonstrated that the induction of carcinomas of the skin of mice by hydrocarbons could be separated into two distinct stages. Only recently, however, have these pioneering experiments been extended to other experimental systems as well as to cancer in the human.

Stages in Carcinogenesis

Some aspects of the basic experiment of Berenblum & Shubik are diagrammed schematically in Figure 1. These workers demonstrated that the stage of carcinogenesis that they termed initiation could be induced by the single application of a relatively small dose of a carcinogen. Following this, if the animals were not subjected to any further form of treatment, no neoplasms resulted (Figure 1, line 1). They further showed that compounds such as urethan when administered by mouth were also capable of initiating the skin (6). Initiation occurred relatively rapidly after the administration of the initiating agent either by mouth or by skin painting, and the change effected in the epidermal cells was not characterized by a specific or even detectable morphological alteration in the structure of the epidermis or dermis. Although not mentioned by Berenblum & Shubik, it is of interest that in the painted area virtually all epidermal cells, numbering in the millions, are



X = Application of Initiator ↓ = Application of Promoter

Figure 1 Schema of the experimental basis for the two-stage process of epidermal carcinogenesis in mice. See text for further details.

exposed to the chemical, but the total number of neoplasms ultimately produced is extremely small (usually less than 20). On the basis of modern concepts of the clonality of neoplasms (7), initiation in mouse skin appears to be an extremely inefficient process as compared with transformation in cell culture (8).

Subsequent to the single administration of the initiating agent, if a second material was repeatedly applied to the initiated area of skin, neoplasms arose in response to the combination of treatments (Figure 1, line 2). This second material was termed a promoting agent, and the process itself, promotion (2, 3). The promoting agent used at first in most experiments was croton oil. This material was highly irritating and stimulated an inflammatory and proliferative response in the epidermis. Most promoting agents induce significant hyperplasia of the skin, although hyperplasia itself without the initiator is insufficient to provoke the neoplastic transformation (9). A number of investigators generally agree that cell replication is necessary for tumor formation but does not appear to be sufficient (10, 11). In the case of urethan initiation of skin, the time at which the croton oil was first administered in relation to the urethan was critical (12). Urethan acted on the skin as a "pure initiator" or incomplete carcinogen in that no neoplasms resulted unless a promoter was subsequently applied. In the case of many polycyclic hydrocarbons and other skin carcinogens such as β -propiolactone and N-methyl-N'-nitro-N-nitrosoguanidine, repeated application of only these materials resulted in skin cancer. Such compounds were termed complete carcinogens, capable of both initiation and promotion. Promoting agents such as croton oil or its purified derivative, tetradecanoyl-phorbol acetate (TPA), when applied to mouse skin in the absence of an initiating agent produced few if any neoplasms (Figure 1, line 5). Although some contest this statement (13), it is interesting to speculate that the few neoplasms produced by continued treatment with promoting agents may result from initiation by ambient environmental initiating agents.

Studies by Tannenbaum (14), Boutwell et al (15), and others demonstrated that the stage of tumor promotion could be modulated by dietary manipulation. Furthermore, studies by Boutwell (16), using graded-dose regimens of croton oil application but a constant total dose of the promoter, demonstrated that tumor promotion was reversible. On the other hand, the process of initiation appears to be irreversible in that the time between initiation of the neoplastic transformation and the beginning of the process of promotion may be prolonged quite extensively (Figure 1, line 3). Although a report by Roe and his associates (17) suggested that an extended time between initiation and promotion resulted in fewer neoplasms, Van Duuren and his associates (18) pointed out that this effect is probably a reflection of the modulation of promotion by the increased age of the

animal. As can be seen from Figure 1, line 4, promotion prior to initiation results in no neoplasms.

Since the original experiments of Berenblum & Shubik (2, 3), other phases or stages of the process of carcinogenesis have also been defined. Boutwell (16) suggested from studies utilizing a combination of croton oil and turpentine that the stage of promotion could be divided into two separate phases, that of *conversion* from an initiated cell to a dormant tumor cell and that of *propagation* of the dormant tumor cell to a grossly visible neoplasm. The latter process was dependent on cellular proliferation, whereas the former was dependent on unspecified intracellular alteration. Foulds (19) formulated an extension of the two-stage concept: he proposed the process of the *progression* of neoplasms from their earliest discernible form, usually a relatively benign lesion, to the highly malignant, rapidly growing neoplasm. The characteristics of tumor progression identified by Foulds included (*a*) the independent progression of tumors, i.e. progression occurs independently in different primary neoplasms within the same host, and (*b*) the independent progression of characters, meaning that the various characteristics of a particular neoplasm progress independently of one another. Such characteristics might include growth rate, invasiveness, metastatic extension, hormone responsiveness, and morphologic appearance. Although it is not always possible to distinguish between the phase of tumor promotion and that of tumor progression, since these two blend with each other in the natural history of many neoplasms, it has been suggested that the latter process is characterized by cellular aneuploidy, whereas tumor promotion may result in neoplasms with minimal or nonexistent identifiable chromosomal abnormalities (20). In addition, a number of neoplasms even shortly after their induction may already represent a fully progressed population of neoplastic cells, yet it is likely that the natural history of most spontaneous neoplasms of man and animals involves all the stages of initiation, promotion, and progression (21). Only within the last decade, however, have stages analogous to those seen in experimental epidermal carcinogenesis been readily demonstrable in tissues other than skin.

Sequential Aspects of the Development of Specific Neoplasms Induced by Chemicals

For almost two decades the stages of initiation and promotion were thought to be applicable only to carcinogenesis of squamous epithelium. Within the past decade, however, a number of systems have been reported in which stages in the pathogenesis of neoplasms can be identified.

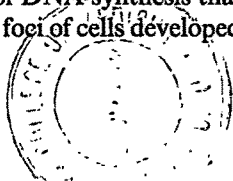
STAGES IN HEPATOCARCINOGENESIS On the theory that hepatocarcinogenesis was a dynamic process occurring in stages, researchers in the

late 1950s and early 1960s made cytologic and cytochemical studies of livers of animals undergoing chemically induced hepatocarcinogenesis. In this way, they hoped to identify cell populations that might be the precursors of hepatocellular carcinomas, and several such studies (22-24) did indeed seem to support such a thesis. Farber and his associates (25, 26) presented both morphological and biochemical evidence that nodules of hepatocytes, termed regenerating or hyperplastic nodules, that resulted from feeding any one of a variety of chemical hepatocarcinogens were the preneoplastic lesions from which malignant hepatocellular carcinomas developed.

Later, Teebor & Becker (27) as well as Farber and his associates (25) demonstrated that the majority of the regenerating or hyperplastic nodules disappeared on cessation of the feeding of the carcinogen. Therefore, despite the more recent suggestion that such lesions should be termed neoplastic nodules (28), it is difficult to understand how the precursor relationship of the nodule to carcinoma can be maintained if the existence of the putative precursor is so transient.

More recently Farber (29) suggested that the cells of the hyperplastic nodule do not disappear when the nodule vanishes from the livers of rats after they are returned to a normal diet, but rather that the cells are "remodeled" and again become part of the normal liver structure. By means of a marker antigen, the PN antigen, found in microsomes of hyperplastic nodules at levels four times greater than that in the surrounding liver (30), Farber and his associates (26) demonstrated that many of the cells of the nodule remained in the remodeled liver and might under appropriate stimuli give rise to neoplasms.

Other investigators have also suggested that cells that are the precursors of malignant hepatocytes may be distinguished by their cytochemistry. Bannasch (31) suggested that "clear" glycogen storage cells evolved through a phase of acidophilia and finally to the basophilic cells generally accepted as probable precursors of hepatic neoplasms. Williams (32) suggested that the inability of cells in foci to accumulate iron in siderotic liver characterized early, possibly preneoplastic lesions. However, earlier studies by Friedrich-Frekxa and his associates (33) demonstrated that in animals given diethylnitrosamine small foci of cells, indistinguishable by standard histopathological techniques from normal liver cells, appeared within six weeks. These cells exhibited a marked deficiency of the enzyme glucose-6-phosphatase and were demonstrable in sections histochemically stained for the enzyme. These investigators theorized that these cells were the precursors of hepatocellular carcinomas. Later, Rabes and his associates (34) showed that islands of tissue deficient in canalicular ATPase, as well as glucose-6-phosphatase, possessed a greater rate of DNA-synthesis than did normal hepatocytes. In a number of such islands, foci of cells developed that



had an increased rate of DNA synthesis and an altered morphology quite comparable to that of hepatocellular carcinomas (35). In these latter studies, Scherer & Emmelot (35) reported that such foci were linearly related to the dose of diethylnitrosamine (up to 30 mg/kg) given within a day after partial hepatectomy. No linear relationship was maintained at doses in excess of 30 mg/kg, but carcinomas developed in such animals, whereas those receiving less than this amount developed only the enzyme-deficient islands or foci of hepatic tissue. At the lower doses the size of the islands increased as the animal aged, but their number did not. These studies strongly supported the precursor relationship of the enzyme-deficient foci to hepatocellular carcinomas and further suggested that diethylnitrosamine in sufficient amounts could "promote" the enzyme-deficient islands to hepatocellular carcinomas.

The first clear evidence that hepatocarcinogenesis could be dissected into stages of its natural history was reported by Peraino and his associates (36), who showed that the feeding of acetylaminofluorene to rats for 3 weeks resulted in an almost negligible incidence of hepatomas 4 months after this regimen. However, if after administration of the carcinogen 0.05% phenobarbital was fed in the diet beginning 1 week after cessation of carcinogen feeding, a tenfold greater number of hepatomas resulted after 4 months. Similar results with the hepatocarcinogens diethylnitrosamine (37) and 3'-methyl-4-dimethylaminoazobenzene (38) have also been reported. In addition, Peraino et al (39) demonstrated that dietary butylated hydroxytoluene, given in a regimen similar to that for phenobarbital, can also enhance or promote acetylaminofluorene-induced liver tumors. In addition, Kimura et al (40) demonstrated that polychlorinated biphenyls were effective promoters in 3'-methyl-4-dimethylaminoazobenzene-induced hepatocarcinogenesis.

More recently Pitot and his associates (41) combined the studies of Scherer & Emmelot (35) with those of Peraino and his associates (36) in order to demonstrate two distinct stages of hepatocarcinogenesis. The administration of a single low dose of diethylnitrosamine following partial hepatectomy (35) resulted in hepatomas only if phenobarbital was administered after initiation by this carcinogen. Although enzyme-altered foci resulted from the partial hepatectomy/diethylnitrosamine treatment, the number of these foci was increased fivefold by subsequent phenobarbital administration. By means of serial frozen sections, these investigators studied the activities of three enzymes in each of the foci and demonstrated that all possible phenotypes were obtained. Farber and his associates (42) earlier showed that deficiencies of cytochrome P₄₅₀ and arylhydrocarbon[benzo-(a)pyrene]hydroxylase were characteristic of many hyperplastic nodules. The deficiencies in these enzymes were utilized in later studies by Solt et al (43) to stimulate the rapid formation of enzyme-altered foci in vivo.

Furthermore, Laishes et al (44) demonstrated that carcinogen resistance could be used to monitor the number of altered cells *in vitro*.

STAGES IN HEPATOCARCINOGENESIS IN THE HUMAN Lesions similar to if not identical with the hyperplastic nodules of the rodent have been identified in the human. The most common of these fall into two groups—the hepatic adenoma and focal nodular hyperplasia (45). These lesions have recently become of considerable interest because of their possible relation to the use of oral contraceptives (46–48). Some of these lesions regressed (49) and a few patients on oral contraceptives exhibited hepatocellular carcinomas (48); these findings suggest that such lesions are analogous to the hyperplastic nodule of the rat in their natural history. Furthermore, the evidence of Peraino and his associates (50) that phenobarbital promoted the spontaneous incidence of hepatomas in mice and that the long-term administration of phenobarbital to rats may result in a late incidence of hepatocellular carcinomas (51) suggests that contraceptive steroids are more likely to be a class of promoting agents than complete or incomplete carcinogens.

STAGES IN CHEMICALLY INDUCED MAMMARY CARCINOGENESIS The existence of stages in mammary carcinogenesis has been established for some years. The existence of the hyperplastic alveolar nodule as an intermediate stage in mammary carcinogenesis was recognized by many investigators (52). Such nodules are probably not the only intermediate stage in mammary carcinogenesis. Medina (53) has suggested that carcinogen-induced mammary ductal hyperplasias can also progress into mammary carcinomas. The hyperplastic alveolar nodule depends for its existence on hormones of the host, especially prolactin (54). Thus prolactin may be considered an endogenous promoting agent for mammary carcinoma in the rodent, although it is clear from the report by Armuth & Berenblum (55) that exogenous agents such as phorbol may act as promoting agents during mammary carcinogenesis by chemicals. Furthermore, removal of prolactin and estrogen from the host internal environment results in the regression of hyperplastic alveolar nodules, which finding suggests that such growths may be analogous to the hyperplastic nodules seen in hepatocarcinogenesis. Whether ductal hyperplasias are comparable to enzyme-altered foci in liver is, however, not apparent.

Although the evidence for a major role of prolactin in human mammary carcinogenesis is not as convincing as that in animals (56), a morphologic investigation suggested that early, possibly preneoplastic lesions in the human breast may be homologous to hyperplastic alveolar nodules and to other lesions found during the development of mammary carcinomas in the rodent (57).

OTHER MODELS OF TWO-STAGE CARCINOGENESIS In addition to skin and liver, a number of other model systems of carcinogenesis exhibit stages of development. Following administration of 1,2-dimethylhydrazine to rats, early changes in the replication and number of intestinal epithelial cells indicate that such foci are probably the precursors of neoplastic lesions of the colon (58). Richards (59) suggested a two-stage process in colon carcinogenesis induced by this compound. Reddy et al (60) presented evidence that bile acids act as promoters of N-methyl-N'-nitro-N-nitrosoguanidine-induced colon cancer in rats. It has long been suspected (61) that colonic polyps are an early stage in the development of malignant neoplasms of the colon in the human.

An excellent example of two-stage carcinogenesis is the induction of bladder cancer in the rat by N-methyl-N-nitrosourea (62). In this study a single dose of the carcinogen instilled in the bladder yielded no neoplasms, but almost all animals fed saccharin for more than 50 weeks during and following such treatment developed bladder carcinomas. The production of hyperplastic nodules in the rat pancreas by the administration of azaserine may also be an example of two-stage carcinogenesis of this organ (63).

Although the production of neoplasia in cell culture is a relatively new phenomenon, studies on the two-stage process of cell transformation in culture have been reported. Mondal & Heidelberger (64) reported that low doses of ultraviolet light given to a mouse cell line in culture resulted in no transformation unless, after irradiation, tetradecanoyl phorbol acetate was added to the cultures. Transformation of rat embryo cells in vitro by polycyclic hydrocarbons was also shown to be promoted by the same agent (65).

MOLECULAR EVENTS IN THE CHEMICAL INITIATION AND PROMOTION OF NEOPLASMS

Knowledge of the natural history of the disease process is incomplete unless the etiology of the disease is understood. Obviously our knowledge of neoplasia is quite incomplete; however, a number of agents that induce neoplasia are known. Without considering the historical aspects of carcinogenesis further, one may summarize our general knowledge in this area by saying that carcinogens—agents that induce the neoplastic transformation in vivo or in vitro—may be classified as physical, chemical, or biologic. In the latter class are viruses and parasites. One may also consider neoplasms resulting directly from congenital or genetic abnormalities as induced by biologic causes. Radiation is the principal physical carcinogen; chemical agents are discussed in greater detail below.

The Chemical Nature of Initiating Agents

Only in this century have we been able to identify specific chemicals that when administered to certain animal species result in the production of neoplasms. The most confusing aspect of these compounds is their varied chemical nature. Table 1, column 1, lists representative examples of several carcinogenic chemicals, some produced naturally and others synthesized for specific uses. The chemical structures of these compounds bear no obvious relation to one another. For years scientists puzzled over this problem, and it was not until the pioneering experiments of the Millers and their associates (4) that a common pattern for all chemical carcinogens became apparent. It is of interest that these studies revealed the metabolism of carcinogens to be the critical factor in their effectiveness, whereas most earlier experimental evidence seemed to indicate that the metabolism of carcinogens destroyed their effectiveness as causative agents of cancer (66).

THE METABOLISM OF CHEMICAL CARCINOGENS AND ITS CONSEQUENCES It has been known for more than two decades (67, 68) that compounds foreign to an organism (including drugs, carcinogens, toxins, etc) are metabolized by enzyme systems located primarily in the endoplasmic reticulum of most cell types, but especially liver cells. In general, these enzymatic reactions involve either oxidation or reduction as well as hydrolysis and/or conjugation with one of several endogenous metabolites such as acetyl or glucuronyl radicals (68). Electronic alterations in the structure of such xenobiotics are catalyzed through a final common pathway to oxygen by the microsomal P₄₅₀ system or, to a lesser extent, by the cytochrome b₅ pathway (20, 69, 71). This review will not consider in detail each

Table 1 Metabolism and metabolic products of some chemical carcinogens^a

Precarcinogen	Proximate form(s)	Ultimate form	Key enzyme reactions
Benzo[a]pyrene	Benzo[a]pyrene-7,8-epoxide	7, 8-Dihydroxy-9, 10-oxy, 7,8,9,10-tetrahydrobenzo[a]pyrene	Aromatic hydroxylation and epoxidation
2-Acetylaminofluorene	N-hydroxy-2-acetylaminofluorene	2-Acetylaminofluorene-N-sulfate	N-hydroxylation and esterification
Aflatoxin B ₁		Aflatoxin B ₁ -2,3-oxide	Epoxidation
Safrole	1'-Hydroxysafrole	1' O-ester of hydroxy safrole	Alkylhydroxylation and esterification
Dimethylnitrosamine	N-Hydroxymethyl,N-methylnitrosamine	Methyl carbonium ion	Alkylhydroxylation
Urethan	Vinyl carbamate	[Ethylene oxide carbamate]	Desaturation and epoxidation

^aSee text for discussion of terms used in column headings. The reader is referred to references 66, 67, and 72 for structural formulas and further details of the enzymatic reactions. The ultimate form of urethan seen in brackets is the proposed form.

of the enzymatic reactions catalyzed by these microsomal enzymes; however, some of the more critical enzyme reactions in the metabolism of carcinogens are listed in the final column of Table 1. All of the enzymatic reactions listed in the table, with the exception of the esterification reactions, involve microsomal oxidative enzymes, especially the P_{450} complex. It should be noted, however, that a number of carcinogenic compounds undergo a series of reactions leading to a highly reactive form. The Millers (72), on the basis of their studies of the metabolism of acetylaminofluorene, showed that the initial metabolic reaction may not lead to the production of a form of the carcinogen directly responsible for the carcinogenic effect. Their studies demonstrated that the first metabolite in the pathway was the hydroxylamine form of acetylaminofluorene. Later studies (73) showed, however, that this compound was nonmutagenic and not more chemically reactive than the parent compound. In their search for further metabolites of this compound, they discovered that esterification of the N-hydroxyl group resulted in an ester that was highly reactive, in the absence of any enzymes, toward a variety of macromolecules including DNA, RNA, and protein. On the basis of these findings they suggested that the initial metabolism of acetylaminofluorene produced a *proximate* form, N-hydroxy-acetylaminofluorene, which was more carcinogenic than the precarcinogen, acetylaminofluorene (74). Further metabolism of the proximate form led to the *ultimate* form of the carcinogen, the N-O- ester, a highly reactive molecule. The table shows the proximate and ultimate forms of several representative chemical carcinogens. It is apparent that virtually all chemical carcinogens that are themselves not ultimate forms, such as the nitrogen mustards (75), undergo reactions to ultimate forms with or without an intermediate proximate form. The enzymatic pathways involved are primarily those associated with microsomal enzymes and the cytochrome P_{450} system.

The Interaction of the Ultimate Forms of Chemical Carcinogens with Cellular Constituents

Since cancer is transmitted in a hereditary manner from cell to cell, present knowledge of molecular biology strongly suggests that the critical site of action of the ultimate forms of chemical carcinogens is DNA. Support for this hypothesis has been increasing steadily as further experimentation in chemical carcinogenesis is carried out. With few exceptions the administration of all carcinogens that have been studied adequately has yielded DNA-, RNA-, and protein-bound derivatives in those tissues in which neoplasms arise (76). The few exceptions to this generalization appear to be compounds such as adriamycin (77), which may be active through tight non-covalent binding to DNA, resulting in an alteration of its replication and/or

transcription; 2,3,7,8-tetrachlorodibenzo-*p*-dioxin, which does not bind to DNA (A. Poland, personal communication), although published studies have indicated its carcinogenicity (78); and plastic or metal films, the implantation of which in subcutaneous sites results in sarcomas (79). This last example of "chemical" carcinogenesis remains an enigma in relation to most theories of carcinogenesis, which are based on the concept of somatic mutation.

Although it is clear that the vast majority of chemical carcinogens are metabolized to an active or ultimate form, if they are not in such a form initially, there is a large class of actual and potential carcinogens that does not fit this pattern. The hormones, both steroid and polypeptide, have been shown to be carcinogenic in a variety of organisms (80). In view of our discussion of the natural history of neoplasms and the fact that under certain circumstances the administration of known promoting agents may induce neoplasms, one may hypothesize that hormones act as promoting agents and that neoplasms resulting from continued hormonal action are initiated by other factors, probably environmental in nature, e. g. chemicals, radiation, and/or biological agents (see above).

The Chemical Nature and Metabolism of Promoting Agents

Like initiating agents, promoting agents also comprise a wide variety of chemical species. For the skin these include phorbol esters, detergents such as polysorbate 80, iodoacetic acid, anthralin, and certain fatty acid esters (11). For the liver, phenobarbital (36), butylated hydroxytoluene (39), and polychlorinated biphenyls (40) are all effective promoting agents. Hormones may also act as promoters (54, 56). As with complete or incomplete carcinogens, an underlying similarity in metabolism and reactivity among promoting agents has been sought. Promotion does not appear to involve any interaction with DNA, since there is no evidence that promoting agents undergo such metabolic or chemical reactions. Exactly what promoting agents do to cells to stimulate the production of tumors from initiated foci is unclear. Boutwell (11) proposed that promoting agents act as gene activators and assembled a substantial amount of information to support his contention. Alternative suggestions include the inhibition of DNA repair (81) by promoting agents, which thus increases the possibility of error during cell replication. Another theory is that promoting agents suppress host immune responsiveness, so that initiated cells can express themselves as neoplasms more efficiently (82).

Although none of these theories has an experimental basis comparable to that on which the mechanism of initiation rests, Boutwell's theory of promoting-agent action fits most of the facts to date. This concept goes far toward reconciling the "carcinogenic" effects of hormones and possibly of

dioxins and related compounds with their lack of direct interaction with the genome of the cell. Many of these compounds activate genes or regulate genetic expression through a receptor mechanism, but do not themselves interact with DNA nor alter it in any way such as by intercalation. Furthermore, evidence has been presented (83) or at least proposed (84) that phorbol esters may interact with specific receptors in or on the cell membrane, in complete analogy to a hormonal-receptor interaction (85, 86).

CONCLUSIONS

Although the etiology of cancer is critical to our final knowledge of the carcinogenic process, a complete understanding of the natural history of neoplasia in organismal, cellular, and molecular terms is our primary hope of understanding and ultimately controlling cancer as a disease. In this brief review we have described our present knowledge of the natural history of the development of neoplasia and some aspects of our knowledge of the molecular events involved in this natural history. Since, on the basis of the prevalence of initiating agents in our environment in the form of air pollution, diet, drugs, radiation, etc, it is quite likely that the initiation of neoplastic cells is a frequent event, a number of oncologists feel that the critical component of the natural history of neoplasia in human cancer is that of promotion. This appears to be the case in many of the more common human neoplasms (87). The study of the basic molecular mechanisms of initiation and promotion may allow us to distinguish clearly between complete carcinogens, incomplete carcinogens, and promoting agents in our environment, and thus provide more rational methods for the protection of the human population from such agents and for the regulation of such agents in our environment.

Literature Cited

1. Shimkin, M. B. 1977. *Contrary to Nature*, pp. 95, 247. Washington, DC: US Dept. Health, Educ., Welfare. 498 pp.
2. Berenblum, I., Shubik, P. 1947. The role of croton oil applications, associated with a single painting of a carcinogen, in tumour induction of the mouse's skin. *Br. J. Cancer* 1:379-91
3. Berenblum, I., Shubik, P. 1949. An experimental study of the initiating stage of carcinogenesis, and a re-examination of the somatic cell mutation theory of cancer. *Br. J. Cancer* 3:109-18
4. Cramer, J. W., Miller, J. A., Miller, E. C. 1960. N-hydroxylation: A new metabolic reaction observed in the rat with the carcinogen, 2-acetylaminofluorene. *J. Biol. Chem.* 235:885-88
5. Conney, A. H., Miller, E. C., Miller, J. A. 1956. The metabolism of methylated aminoazo dyes. V. Evidence for induction of enzyme synthesis in the rat by 3-methylcholanthrene. *Cancer Res.* 16:450-59
6. Berenblum, I., Haran-Ghera, N. 1957. A quantitative study of the systemic initiating action of urethane (ethyl carbamate) in mouse skin carcinogenesis. *Br. J. Cancer* 11:77-84
7. Fialkow, P. J. 1976. Clonal origin of human tumors. *Biochim. Biophys. Acta* 458:283-321
8. Kakunaga, T. 1975. The role of cell division in the malignant transformation of mouse cells treated with 3-methylcholanthrene. *Cancer Res.* 35:1637-42

9. Shubik, P. 1950. Studies on the promoting phase in the stages of carcinogenesis in mice, rabbits, and guinea pigs. *Cancer Res.* 10:13-19
10. Salaman, M. H., Roe, F. J. C. 1964. Cocarcinogenesis. *Br. Med. Bull.* 20: 139-44
11. Boutwell, R. K. 1974. The function and mechanism of promoters of carcinogenesis. *Crit. Rev. Toxicol.* 2:419-43
12. Yuspa, S. H., Hennings, H., Saffiotti, U. 1976. Cutaneous chemical carcinogenesis: Past, present, and future. *J. Invest. Dermatol.* 67:199-208
13. Roe, F. J. C., Clack, J. 1964. Two-stage carcinogenesis: The effect of length of promoting treatment on the yield of benign and malignant tumors. *Br. J. Cancer* 17:596-604
14. Tannenbaum, A. 1944. The dependence of the genesis of induced skin tumors on the fat content of the diet during different stages of carcinogenesis. *Cancer Res.* 4:683-87
15. Boutwell, R. K., Brush, M. K., Rusch, H. P. 1949. The influence of vitamins of the B complex on the induction of epithelial tumors in mice. *Cancer Res.* 9:747-52
16. Boutwell, R. K. 1964. Some biological aspects of skin carcinogenesis. *Progr. Exp. Tumor Res.* 4:207-50
17. Roe, F. J. C., Carter, R. C., Mitchley, B. C. V., Peto, R., Hecker, E. 1972. On the persistence of tumor initiation and the acceleration of tumor progression in mouse skin tumorigenesis. *Int. J. Cancer* 9:264-73
18. Van Duuren, B. L., Sivak, A., Katz, C., Seidman, I., Melchionne, S. 1975. The effect of aging and interval between primary and secondary treatment in two-stage carcinogenesis on mouse skin. *Cancer Res.* 35:502-5
19. Foulds, L. 1965. Multiple etiologic factors in neoplastic development. *Cancer Res.* 25:1339-47
20. Pitot, H. C. 1978. *Fundamentals of Oncology*, p. 108. New York: Dekker. 192 pp.
21. Pitot, H. C. 1977. The natural history of neoplasia—newer insights into an old problem. *Am. J. Pathol.* 89:402-11
22. Weiler, E. 1956. Die Änderung der serologischen Spezifität von Leberzellen der Ratte während der Cancerogenese durch *p*-Dimethylaminobenzol. *Z. Naturforsch.* 11b:31-38
23. Hughes, P. E. 1958. The significance of staining reactions of preneoplastic rat liver with fluorescein-globulin complexes. *Cancer Res.* 18:426-32
24. Goldfarb, S., Zak, F. G. 1961. Role of injury and hyperplasia in the induction of hepatocellular carcinoma. *J. Am. Med. Assoc.* 178:729-31
25. Epstein, S., Ito, N., Merkow, L., Farber, E. 1967. Cellular analysis of liver carcinogenesis: The induction of large hyperplastic nodules in the liver with 2-fluorenylacetamide or ethionine and some aspects of their morphology and glycogen metabolism. *Cancer Res.* 27:1702-11
26. Farber, E. 1973. Hyperplastic liver nodules. *Methods Cancer Res.* 7:345-75
27. Teebor, G. W., Becker, F. F. 1977. Regression and persistence of hyperplastic nodules induced by N-2-fluorenylacetamide and their relationship to hepatocarcinogenesis. *Cancer Res.* 31:1-3
28. Squire, R. A., Levitt, M. H. 1975. Report of a workshop on classification of specific hepatocellular lesions in rats. *Cancer Res.* 35:3214-23
29. Farber, E. 1976. On the pathogenesis of experimental hepatocellular carcinoma. In *Hepatocellular Carcinoma*, ed. K. Okuda, R. L. Peters, pp. 3-22. New York: Wiley. 499 pp.
30. Griffin, M. J., Kizer, D. E. 1978. Purification and quantitation of preneoplastic antigen from hyperplastic nodules and normal liver. *Cancer Res.* 38:1136-41
31. Bannasch, P. 1976. Cytology and cytogenesis of neoplastic (hyperplastic) hepatic nodules. *Cancer Res.* 36: 2555-62
32. Williams, G. M. 1976. Functional markers and growth behavior of preneoplastic hepatocytes. *Cancer Res.* 36:2540-43
33. Friedrich-Freksa, H., Gössner, W., Börner, P. 1969. Histochemische Untersuchungen der Cancerogenese in der Rattenleber nach Dauergaben von Diäthylnitrosamin. *Z. Krebsforsch.* 72: 226-39
34. Rabes, H. M., Schultz, P., Jantsch, B. 1972. Growth kinetics of diethylnitrosamine-induced, enzyme-deficient "preneoplastic" liver cell populations *in vivo* and *in vitro*. *Cancer Res.* 32:2577-89
35. Scherer, E., Emmelot, P. 1975. Kinetics of induction and growth of precancerous liver-cell foci, and liver tumor formation by diethylnitrosamine in the rat. *Eur. J. Cancer* 11:689-701
36. Peraino, C., Fry, R. J. M., Staffeldt, E., Kisieleski, W. E. 1973. Effects of varying the exposure to phenobarbital on its enhancement of 2-acetylaminofluorene-

- induced hepatic tumorigenesis in the rat. *Cancer Res.* 33:2701-5
37. Weisburger, J. H., Madison, R. N., Ward, G. M., Viguera, C., Weisburger, E. 1975. Modification of diethylnitrosamine liver carcinogenesis with phenobarbital but not with immunosuppression. *J. Natl. Cancer Inst.* 54:1185-88
 38. Kitagawa, T., Sugano, H. 1977. Enhancement of azo-dye hepatocarcinogenesis with dietary phenobarbital in rats. *Gann* 68:255-56
 39. Peraino, C., Fry, R. J. M., Staffeldt, E., Christopher, J. P. 1977. Enhancing effects of phenobarbitone and butylated hydroxytoluene on 2-acetylaminofluorene-induced hepatic tumorigenesis in the rat. *Food Cosmet. Toxicol.* 15:93-96
 40. Kimura, N. T., Kanematsu, T., Baba, T. 1976. Polychlorinated biphenyl(s) as a promoter in experimental hepatocarcinogenesis in rats. *Z. Krebsforsch.* 87:257-66
 41. Pitot, H. C., Barsness, L., Goldsworthy, T., Kitagawa, T. 1978. Biochemical characterization of stages of hepatocarcinogenesis after a single dose of diethylnitrosamine. *Nature* 271:456-58
 42. Cameron, R., Sweeney, G. D., Jones, K., Lee, G., Farber, E. 1976. A relative deficiency of cytochrome P-450 and aryl hydrocarbon [benzo(a)pyrene] hydroxylase in hyperplastic nodules induced by 2-acetylaminofluorene in rat liver. *Cancer Res.* 36:3888-93
 43. Solt, D. B., Medline, A., Farber, E. 1977. Rapid emergence of carcinogen-induced hyperplastic lesions in a new model for the sequential analysis of liver carcinogenesis. *Am. J. Pathol.* 88:595-618
 44. Laishes, B. A., Roberts, E., Farber, E. 1978. *In vitro* measurement of carcinogen-resistant liver cells during hepatocarcinogenesis. *Int. J. Cancer* 21:186-93
 45. Sorensen, T. I. A., Baden, H. 1975. Benign hepatocellular tumors. *Scand. J. Gastroenterol.* 10:113-19
 46. Goldfarb, S. 1976. Sex hormones and hepatic neoplasia. *Cancer Res.* 36:2584-88
 47. Klatskin, G. 1977. Hepatic tumors: Possible relationship to oral contraceptives. *Gastroenterology* 73:386-94
 48. Christopherson, W. M., Mays, E. T. 1977. Liver tumors and contraceptive steroids: Experience with the first hundred registry patients. *J. Natl. Cancer Inst.* 58:167-71
 49. Edmondson, H. A., Reynolds, T. B., Henderson, B., Benton, B. 1977. Regression of liver cell adenomas associated with oral contraceptives. *Ann. Intern. Med.* 86:180-82
 50. Peraino, C., Fry, R. J. M., Steffeldt, E. 1973. Enhancement of spontaneous hepatic tumorigenesis in C3H mice by dietary phenobarbital. *J. Natl. Cancer Inst.* 51:1349-50
 51. Rossi, L., Ravera, M., Repetti, G., Santì, L. 1977. Long-term administration of DDT or phenobarbital-Na in Wistar rats. *Int. J. Cancer* 19:179-85
 52. Nandi, S., McGrath, C. M. 1973. Mammary neoplasia in mice. *Adv. Cancer Res.* 17:353-414
 53. Medina, D. 1976. Preneoplastic lesions in murine mammary cancer. *Cancer Res.* 36:2589-95
 54. Welsch, C. W., Nagasawa, H. 1977. Prolactin and murine mammary tumorigenesis: A review. *Cancer Res.* 37:951-63
 55. Armuth, V., Berenblum, I. 1974. Promotion of mammary carcinogenesis and leukemogenic action by phorbol in virgin female Wistar rats. *Cancer Res.* 34:2704-7
 56. Smithline, F., Sherman, L., Kolodny, H. D. 1975. Prolactin and breast carcinoma. *N. Engl. J. Med.* 292:784-92
 57. Jensen, H. M., Rice, J. R., Wellings, S. R. 1976. Preneoplastic lesions in the human breast. *Science* 191:295-97
 58. Lipkin, M. 1974. Proliferative changes in the colon. *Am. J. Dig. Dis.* 19:1029-32
 59. Richards, T. C. 1977. Early changes in the dynamics of crypt cell populations in mouse colon following administration of 1,2-dimethylhydrazine. *Cancer Res.* 37:1680-85
 60. Reddy, B. S., Watanabe, K., Weisburger, J. H., Wynder, E. L. 1977. Promoting effect of bile acids in colon carcinogenesis in germ-free and conventional F344 rats. *Cancer Res.* 37:3238-42
 61. Kozuka, S., Taki, T., Kubota, K., Yokoyama, Y. 1976. Simultaneous transition of multiple polyps into carcinoma in the large intestine. *Dis. Colon Rectum* 19:655-59
 62. Hicks, R. M., Wakefield, J. St. J., Chowanec, J. 1973. Co-carcinogenic action of saccharin in the chemical induction of bladder cancer. *Nature* 243:347-49
 63. Longnecker, D. S., French, J., Hyde, E., Lilja, H. S., Yager, J. D. Jr. 1977. Effect of age on nodule induction by azaserine and DNA synthesis in rat pancreas. *J. Natl. Cancer Inst.* 58:1769-75

64. Mondal, S., Heidelberger, C. 1976. Transformation of C3H/10T1/2 CL8 mouse embryo fibroblasts by ultraviolet irradiation and a phorbol ester. *Nature* 260:710-11
65. Lasne, C., Gentil, A., Chouroulinkov, I. 1977. Two-stage carcinogenesis with rat embryo cells in tissue culture. *Br. J. Cancer* 35:722-29
66. Miller, J. A., Miller, E. C. 1965. Natural and synthetic chemical carcinogens in the etiology of cancer. *Cancer Res.* 25:1292-1304
67. Hucker, H. B. 1973. Intermediates in drug metabolism reactions. *Drug Metab. Rev.* 2:33-56
68. Drayer, D. D. 1974. Pathways of drug metabolism in man. *Med. Clin. North Am.* 58:927-44
69. Gillette, J. R., Davis, D. C., Sasame, H. A. 1972. Cytochrome P-450 and its role in drug metabolism. *Ann. Rev. Pharmacol.* 12:57-84
70. Lu, A. Y. H., Levin, W. 1974. The resolution and reconstitution of the liver microsomal hydroxylation system. *Biochim. Biophys. Acta* 344:205-40
71. Schenkman, J. D., Jansson, I., Robie-Suh, K. N. 1976. The many roles of cytochrome *b₅* in hepatic microsomes. *Life Sci.* 19:611-24
72. Miller, E. C., Miller, J. A. 1966. Mechanisms of chemical carcinogenesis: Nature of proximate carcinogens and other actions with macromolecules. *Pharmacol. Rev.* 18:805-38
73. Maher, V. M., Miller, E. C., Miller, J. A., Szybalski, W. 1968. Mutations and decreases in density of transforming DNA produced by derivatives of the carcinogens 2-acetylaminofluorene and *N*-methyl-4-aminoazobenzene. *Mol. Pharmacol.* 4:411-26
74. Miller, E. C., Miller, J. A., Enomoto, M. 1964. The comparative carcinogenicities of 2-acetylaminofluorene and its *N*-hydroxy metabolite in mice, hamsters, and guinea pigs. *Cancer Res.* 24:2018-32
75. Lawley, P. D. 1976. Carcinogenesis by alkylating agents. In *Chemical Carcinogens*, ed. C. E. Searle, Am. Chem. Soc. Monogr. 173, pp. 83-244. Washington, DC: Am. Chem. Soc. 788 pp.
76. Miller, E. C. 1978. Some current perspectives on chemical carcinogenesis in humans and experimental animals: Presidential address. *Cancer Res.* 38:1479-96
77. Marquardt, H., Baker, S., Grab, D. 1977. Oncogenic and mutagenic activity of adriamycin decreased by microsomal metabolism. *Proc. Am. Assoc. Cancer Res.* 18:13 (Abstr. No. 50).
78. Van Miller, J. P., Lalich, J. J., Allen, J. R. 1977. Increased incidence of neoplasms in rats exposed to low levels of 2,3,7,8-tetrachlorodibenzo-*p*-dioxin. *Chemosphere* 6:537-44
79. Brand, K. G., Buoen, L. C., Johnson, K. H., Brand, I. 1975. Etiological factors, stages, and the role of the foreign body in foreign body tumorigenesis: A review. *Cancer Res.* 35:279-86
80. Furth, J. 1975. Hormones as etiological agents in neoplasia. In *Cancer—A Comprehensive Treatise*, ed. F. F. Becker, 1:75-120. New York: Plenum. 524 pp.
81. Gaudin, D., Gregg, S., Yielding, K. L. 1971. DNA repair inhibition: A possible mechanism of action of co-carcinogens. *Biochem. Biophys. Res. Commun.* 45:630-36
82. Schinitzky, M. R., Hyman, L. R., Blazkovec, A. A., Burkholder, P. M. 1973. *Bacillus Calmette-Guerin* vaccination and skin tumor promotion with croton oil in mice. *Cancer Res.* 33:659-63
83. Slaga, T. J., Scribner, J. D., Rice, J. M., Das, S. B., Thompson, S. 1974. Inhibition by dexamethasone of intracellular binding of phorbol esters to mouse skin. *J. Natl. Cancer Inst.* 52:1611-18
84. Rohrschneider, L. R., Boutwell, R. K. 1973. Phorbol esters, fatty acids and tumour promotion. *Nature New Biol.* 243:212-13
85. Liao, S. 1975. Cellular receptors and mechanisms of action of steroid hormones. *Int. Rev. Cytol.* 41:87-172
86. Ryan, R. J., Lee, C. Y. 1976. The role of membrane-bound receptors. *Biol. Reprod.* 14:16-29
87. Wynder, E. L., Hoffmann, D., McCoy, G. D., Cohen, L. A., Reddy, B. S. 1978. Tumor promotion and cocarcinogenesis as related to man and his environment. In *Mechanisms of Tumor Promotion and Cocarcinogenesis*, ed. T. J. Slaga, A. Sivak, R. K. Boutwell, 2:59-77. New York: Raven. 588 pp.

ANGIOGRAPHIC MANAGEMENT ♦7305 OF GASTROINTESTINAL BLEEDING¹

Saadoon Kadir, M.D.² and Christos A. Athanasoulis, M.D.

Departments of Radiology, Massachusetts General Hospital
and Harvard Medical School, Boston, Massachusetts 02114

Localization of gastrointestinal bleeding sites and the control of bleeding with vasopressin or transcatheter embolization are new applications of selective visceral angiography. In the majority of patients, gastrointestinal bleeding stops on conservative therapy and therefore the number of patients requiring diagnostic and therapeutic angiography is relatively small. However, when conservative measures fail, angiographic methods may be applied in order to prevent a major surgical intervention.

In general, endoscopy should precede angiography, as endoscopic localization of the bleeding site can facilitate the angiographic procedure. On the other hand, barium studies are of little help in the evaluation of the acutely bleeding patient and may unnecessarily delay both angiography and endoscopy.

METHODS

Arteriography is performed via the femoral route using the techniques described by Seldinger (1) and Ödman (2). The preceding endoscopy or clinical presentation determines the vessels to be selectively catheterized. For upper GI bleeding from the stomach we proceed directly to the selective catheterization of the left gastric artery and use the loop catheter technique (3). For duodenal bleeding a celiac arteriogram is performed first. For

¹This work was supported in part by US Public Health Grants BRSG-0548613 and 1R01 AM 19811-02.

²Present Address: Department of Radiology, Vanderbilt University Hospital, Nashville, Tennessee 37232.

colonic bleeding the superior mesenteric arteriogram is followed by an inferior mesenteric arteriogram. This sequence is reversed if clinical or endoscopic evidence suggests a left colonic source for the bleeding. The bleeding site is identified with the demonstration of contrast medium extravasation on serial films. Once the bleeding site has been localized it can be treated with intraarterial or systemic infusion of vasopressin or transcatheter vessel occlusion.

VASOPRESSIN

Vasopressin, which was used as early as 1956 for the management of bleeding esophageal varices, has a dual action causing constriction of the smooth musculature of the bowel wall and blood vessels (4, 5).

The choice of intravenous vs intraarterial vasopressin is determined by the location and type of bleeding. Our clinical experience shows that mucosal, variceal, and occasionally low grade colonic bleeding may respond to infusions of intravenous vasopressin. Otherwise the intraarterial route is used.

Once the bleeding site has been localized, vasopressin is infused for 20 minutes. A control arteriogram is then performed to evaluate the angiographic results. If the bleeding has been controlled, the patient is managed according to the following infusion schedule: 0.2 U/min for 12–24 hr; 0.1 U/min for 12–24 hr; D5W or NS at 20–30 ml/hr to keep the catheter open. Meanwhile the patient is observed for signs of rebleeding (hematocrit, nasogastric aspirate, central venous pressure, and transfusion requirement). If bleeding has stopped the catheter is removed 6–12 hours later. If bleeding persists transcatheter occlusion is considered.

Intraarterial vasopressin infusions are performed with a constant infusion pump. If systemic vasopressin is used, the angiographic catheter is removed after completion of the diagnostic procedure.

TRANSCATHETER OCCLUSION

The materials most commonly used for transcatheter occlusion of the bleeding vessel are blood clot and Gelfoam.[®] Ivalon[®], a compressible plastic foam of polyvinyl alcohol, and isobutyl-2-cyanoacrylate find occasional use in situations where a permanent occlusion is desirable. The latter is not yet available for general use. Detachable small balloons attached to angiographic catheters have recently been introduced for the control of bleeding.

The embolization procedure is preformed through the arteriography catheter. When superselective embolization becomes necessary a coaxial system using a 3-French, straight, Teflon[®] catheter is used. Gelfoam is cut

into 2-mm square pieces. Prior to the embolization it is soaked in saline and, with a tuberculin syringe, the Gelfoam is loaded into the catheter and flushed with 1–2 ml of saline. The number of Gelfoam pieces necessary to occlude a bleeding vessel depends upon its size and location. For visceral arteries usually 2–5 plugs are sufficient. A similar technique is used for blood clot and Ivalon.

For the delivery of isobutyl-2-cyanoacrylate, a coaxial system is used. A 3-French, straight, Teflon catheter is introduced through the arteriography catheter. Care must be taken to use nonionic solutions (5% dextrose) to flush the catheter, as the tissue adhesive polymerizes almost instantaneously upon contact with an ionic surface, which thus leads to the incorporation of the catheter in the polymer.

ESOPHAGEAL BLEEDING

Bleeding from the lower esophagus can be demonstrated by selective left gastric arteriography. This includes bleeding from Mallory-Weiss tears, esophageal tumors, reflux esophagitis, or trauma following instrumentation (Figure 1). Selective infusion of vasopressin in the left gastric artery is successful in controlling the bleeding in most cases.

GASTRIC BLEEDING

Bleeding from peptic ulcers, which can often be massive, is best demonstrated by a left gastric arteriogram. Intraarterial infusion of vasopressin can successfully stop the bleeding in 65–70% of patients (6, 7). In those patients where intraarterial vasopressin fails to control the bleeding, transcatheter embolization is considered (Figure 2). Mucosal hemorrhage due to gastritis or superficial ulcerations can be controlled in 84% of patients with intraarterial vasopressin. Recurrent bleeding is observed in 16% (8).

DUODENAL BLEEDING

Bleeding from a duodenal ulcer is demonstrated on a celiac arteriogram. As vasopressin infusions are unsuccessful in controlling duodenal hemorrhage in about 50% of these patients, transcatheter occlusion appears to be the angiographic method of choice for controlling the hemorrhage (Figure 3).

The ineffectiveness of intraarterial vasopressin in controlling peptic ulcer bleeding in the duodenum and stomach is due to the presence of chronic inflammatory changes at the base of the ulcer that prevent adequate vasoconstrictive response to vasopressin. Furthermore the dual blood supply to the duodenum may be responsible for the vasopressin failures.

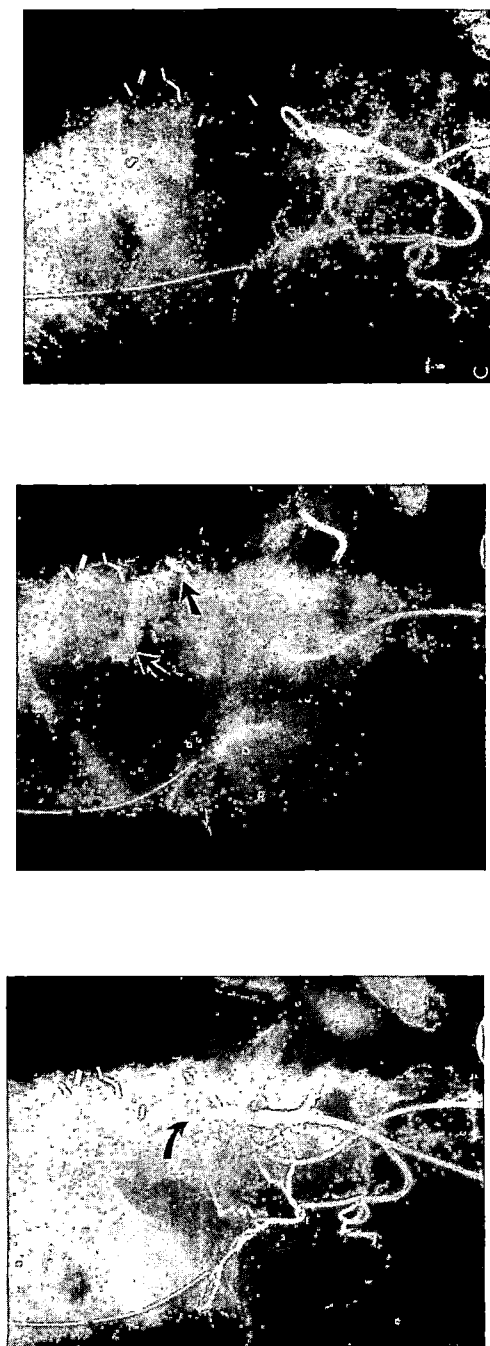


Figure 1 Distal esophageal bleeding due to reflux esophagitis. (A) The left gastric artery arises from the left hepatic artery. Contrast extravasation is seen from the esophageal branches (*arrow*). (B) Late arterial phase shows further extravasation (*arrows*). (C) Following intraarterial infusion of vasopressin the bleeding stopped.

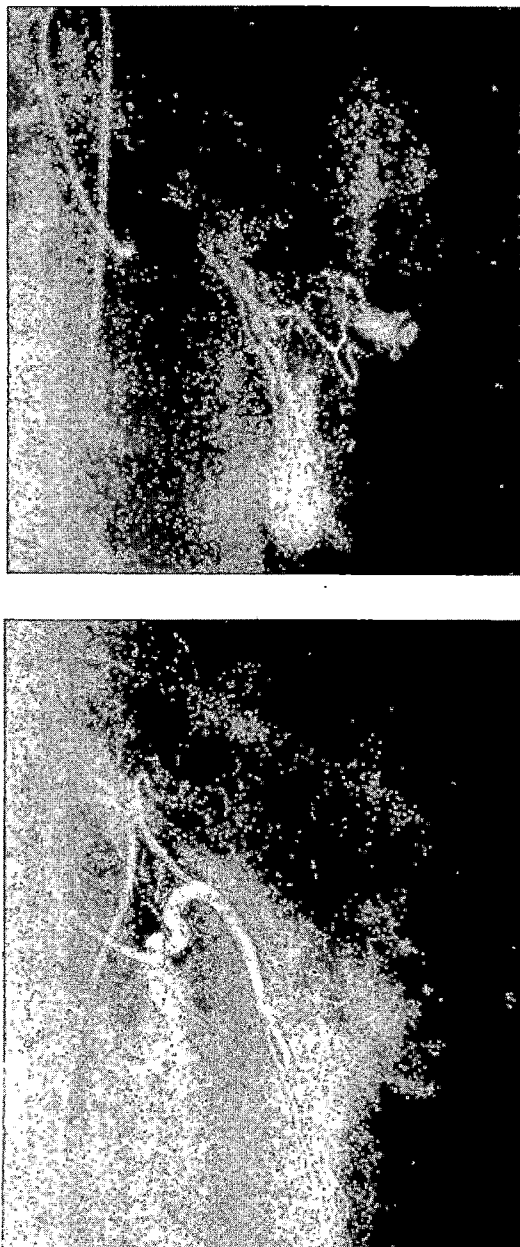


Figure 2 Gastric stress ulcer bleeding. (A) Left gastric arteriogram shows contrast extravasation. (B) After transcatheter embolization with Gelfoam, bleeding that could not be controlled by intraarterial vasopressin stopped.

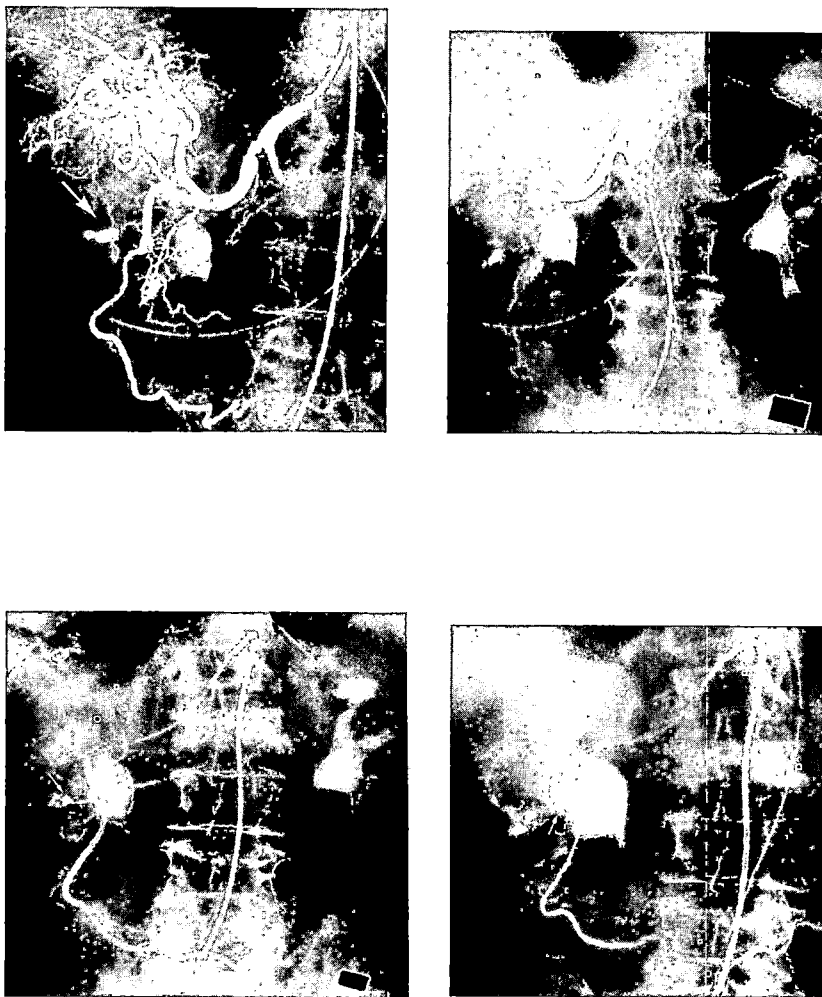


Figure 3 Bleeding from a duodenal ulcer. (A) Common hepatic arteriogram demonstrates extravasation of contrast medium (*arrow*). (B) Control arteriogram following 20-min infusion of 0.3-U/min vasopressin in the common hepatic artery. The extravasation persists. (C) Pre-embolization gastroduodenal arteriogram using a coaxial system shows persistent bleeding (*arrow*). (D) After transcatheter occlusion with Gelfoam, the arterial branches to the bleeding ulcer have been occluded (*arrow*). No further contrast medium extravasation is seen.

PORTAL HYPERTENSION

Bleeding from varices is difficult to demonstrate by mesenteric arteriography because by the time the contrast medium reaches the veins, it is diluted. The diagnosis is made by endoscopy. However, in a patient with upper gastrointestinal bleeding and portal hypertension with varices demonstrated by endoscopy and angiography, the source of bleeding may be a peptic ulcer or gastritis (9). Selective arteriography is able to determine the site of bleeding and delineate the portal venous anatomy if a shunt procedure is considered. For variceal bleeding, once other sources of bleeding have been excluded, intravenous vasopressin is infused. This decreases portal venous pressure by 35–50%, thereby decompressing the varices and facilitating thrombosis (9). Recent clinical and laboratory experience shows that both superior mesenteric artery and intravenous vasopressin in equivalent dose rates reduce portal venous pressure to a comparable degree (10).

Definitive therapy for variceal bleeding is a portosystemic shunt. As a temporizing maneuver to prepare the patient for definitive treatment, transhepatic obliteration can be used. After percutaneous transhepatic catheterization of the coronary vein, the varices are embolized with Gelfoam or isobutyl-2-cyanoacrylate (Figure 4). Before removing the transhepatic catheter the intrahepatic tract is embolized with Gelfoam. This reduces the incidence of complications, such as hemoperitoneum and peritoneal irritation. In experienced hands this procedure is relatively safe and offers an attractive alternative to emergency surgery that has a high mortality.

SMALL BOWEL

Common sources of bleeding from the small bowel are tumors, vascular malformations, ulcerations, and occasionally diverticula (Figure 5) (9). Selective arteriography may localize the bleeding, which can then be controlled effectively with intraarterial vasopressin.

COLONORECTAL BLEEDING

Massive colonic bleeding is most commonly due to diverticular bleeding or ulcerative colitis (9). Since up to 75% of the bleeding diverticula are located in the right hemicolon, the superior mesenteric arteriogram is performed first (Figure 6) (11). If no contrast extravasation is demonstrated or prior endoscopy shows a left-sided lesion, an inferior mesenteric arteriogram is performed first. If no bleeding is demonstrated, celiac arteriography should be done to exclude a hitherto unrecognized source of bleeding.

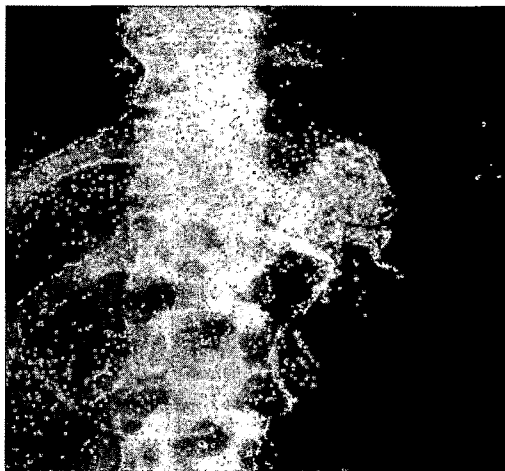
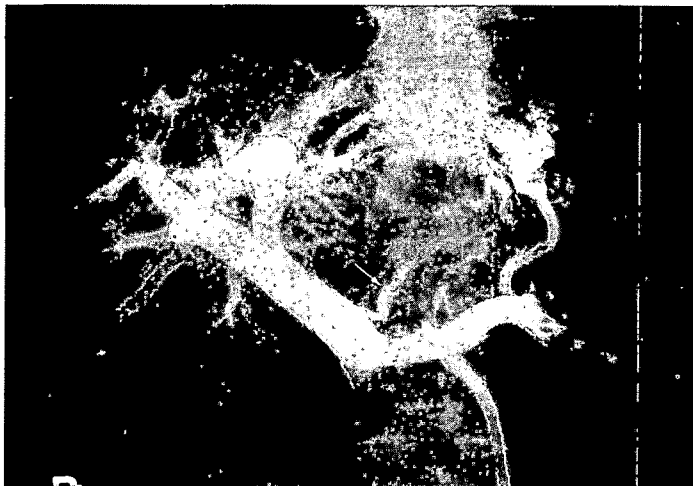


Figure 4 Transhepatic obliteration of gastroesophageal varices for endoscopy-proven variceal bleeding. (A) Contrast injection in the splenic vein: There is fugal flow into the inferior mesenteric vein and opacification of short gastric varices and right gastric vein (*arrow*). (B) Catheterization of the short gastric varices using a coaxial system. (C) Selective contrast injection into the coronary varices. (D) Splenic venogram following embolization of the short gastric and coronary varices. The bleeding stopped and the patient was discharged.



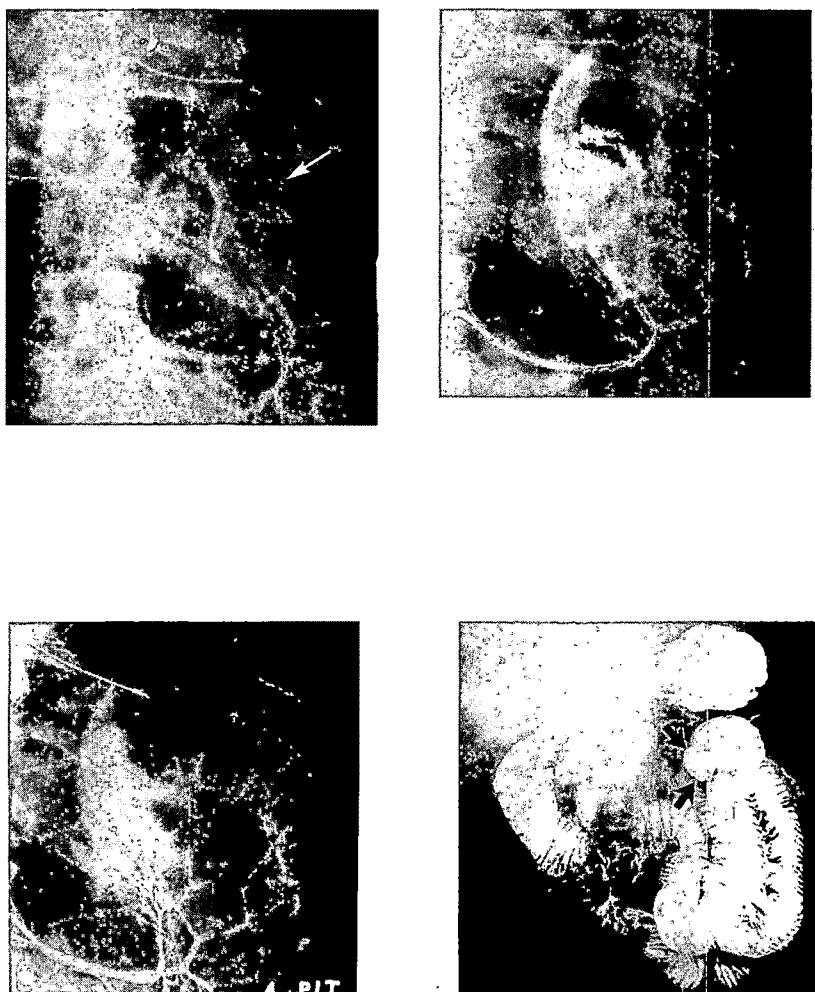


Figure 5 Bleeding from a jejunal diverticulum. (A) Superior mesenteric arteriogram shows extravasation of contrast medium from a branch of the first jejunal artery (*arrow*). (B) Selective contrast injection into the first jejunal artery redemonstrates the bleeding site. (C) The bleeding stopped temporarily following intraarterial infusion of vasopressin. At surgery the bleeding jejunal diverticulum was resected. (D) Upper gastrointestinal barium examination performed several days prior to the surgery shows a large proximal jejunal diverticulum (*arrows*).

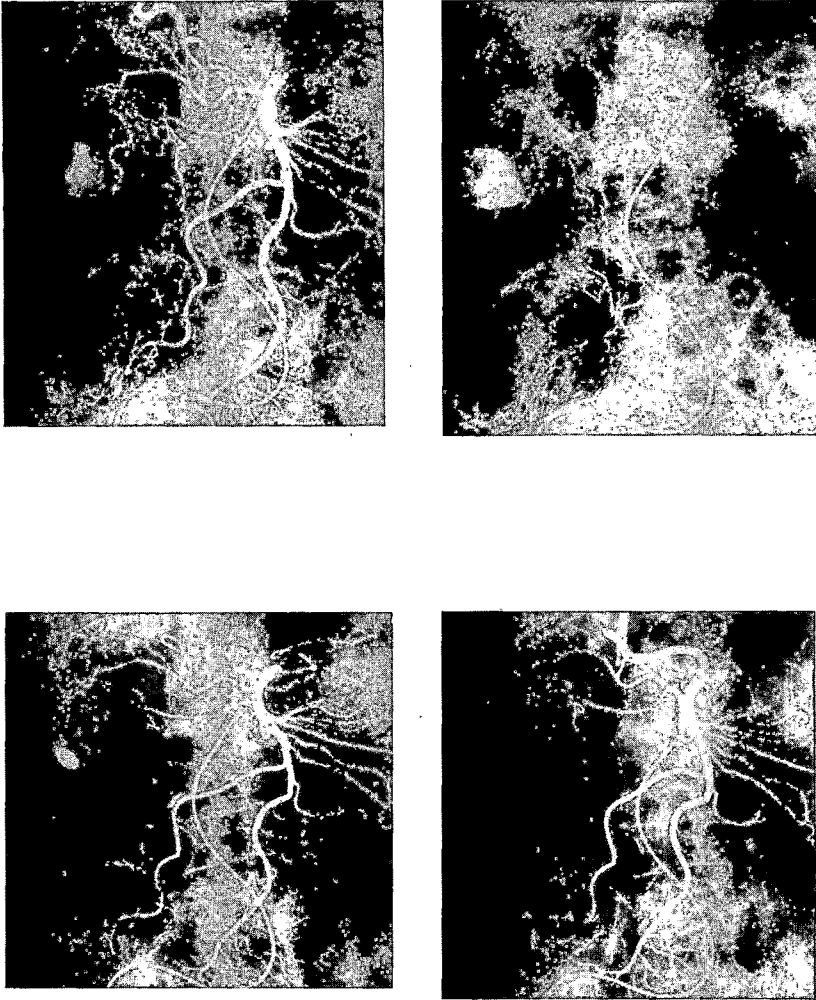


Figure 6 Bleeding from a right colon diverticulum. (A) Early arterial phase of a superior mesenteric arteriogram demonstrates extravasation of contrast in the ascending colon. (B) Late arterial phase from the same arteriogram shows further extravasation. (C) Intravenous vasopressin failed to control the bleeding. (D) The bleeding stopped following infusion of 0.2-U/min vasopressin via the superior mesenteric artery.

Intraarterial vasopressin can successfully control acute colonorectal bleeding in up to 90% of patients. Recurrent bleeding is observed in around 25% of patients generally within the first 24 hours (12). However, localization of the bleeding site helps to modify the operative procedure, thus lowering operative mortality.

Lower intestinal bleeding can also be due to mesenteric varices that may develop after abdominal surgery or as a result of mesenteric vein obstruction caused by pancreatic carcinoma, mesenteric panniculitis, carcinoid, and portal hypertension (Figure 7). Although these can be demonstrated on the venous phase of the superior mesenteric arteriogram, for better visualization transhepatic portography may be used. Here the management is primarily surgical but intravenous vasopressin can decompress the varices and thus lead to temporary cessation of the bleeding.

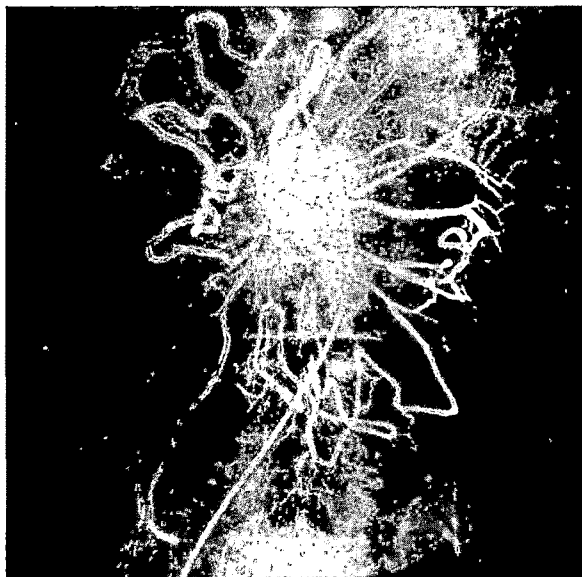
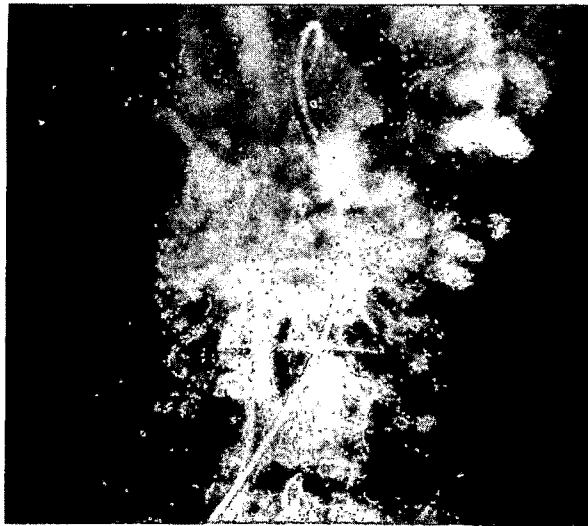


Figure 7 Lower intestinal bleeding in a patient with carcinoid of the mesentery. (A) Arterial phase of the superior mesenteric arteriogram: There is a typical fixed, stellate configuration of the arteries with encasement and occlusion of several branches. (B) Superior mesenteric arteriogram: Close-up view in the right posterior oblique projection shows the arterial occlusions and encasement in greater detail. (C) There is a similar stellate configuration of the veins and occlusion of the superior mesenteric vein with visualization of multiple dilated collaterals.



COLONIC ARTERIOVENOUS MALFORMATIONS

Arteriovenous malformations of the colon may cause acute rectal bleeding. The diagnosis is made by selective arteriography of the superior and inferior mesenteric arteries. If the arteriovenous malformation is located in the sigmoid colon or rectum, bilateral internal iliac arteriograms are also performed. As these arteriovenous malformations do not respond to intraarterial vasopressin, the angiographic treatment is transcatheter occlusion (Figure 8).



Figure 8 Acute lower intestinal bleeding from a sigmoid arteriovenous malformation in a 21-year-old female. (A) The inferior mesenteric arteriogram demonstrates a large arteriovenous malformation. (B) After embolization with Gelfoam through a coaxial system, the bleeding stopped but recurred 2½ months later.

ANGIODYSPLASIA

In patients 55 years and older, intermittent low grade, and occasionally massive, lower GI bleeding may be due to angiodysplasia of the right colon or cecum. The characteristic angiographic appearance of angiodysplasia is (Figure 9): (*a*) early draining vein; (*b*) persistent and dense opacification of the vein; (*c*) vascular tuft on the antimesenteric border; and (*d*) on occasion contrast extravasation.

The lesions are submucosal and can on occasion be identified by the experienced colonoscopist. Surgical exploration and barium examination are usually negative. However, a barium examination is essential to exclude coexisting lesions, which have been observed in 22% of patients (13). The treatment of choice for angiodysplasia is a right hemicolectomy. More recently, electrocoagulation has been used with success.

POSTOPERATIVE BLEEDING

Angiography has made an important contribution to the diagnosis and therapy of postoperative bleeding (14). Such bleeding can occur in the immediate postoperative period as a result of a slipped ligature or inadequate ligation of bleeding vessels, or it may be seen several days later following mucosal sloughing.

Selective arteriography of the major visceral artery supplying the operated area will demonstrate the source of bleeding. Treatment includes intravenous or intraarterial infusion of vasopressin or transcatheter occlusion (Figure 10). Occasionally the bleeding may be only indirectly related to the surgery as in stress ulcer bleeding. Here the appropriate vessel is catheterized and treated with vasopressin infusions.

COMPLICATIONS

Complications of angiographic management of gastrointestinal bleeding are similar to those observed in patients undergoing diagnostic arteriography (15). Vasopressin can be infused for several days without any serious side effects. However, its use under general anesthesia and with ganglion blockers should be cautious (5). Anginal pains have been observed in patients with ischemic heart disease. Abdominal cramps, nausea, and vomiting are seen occasionally. Water and electrolyte imbalance with a temporary decrease in urine output, which is a result of the antidiuretic hormone effect of vasopressin, can be managed with diuretics.

Following aberrant vasopressin infusion, premature ventricular contractions, bradycardia, hematuria, and peripheral vasoconstriction have

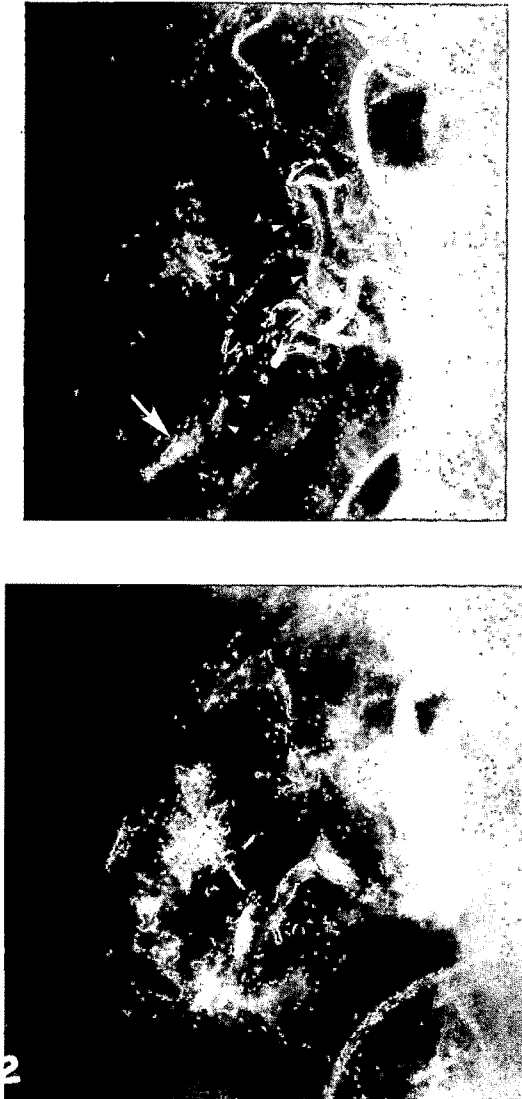
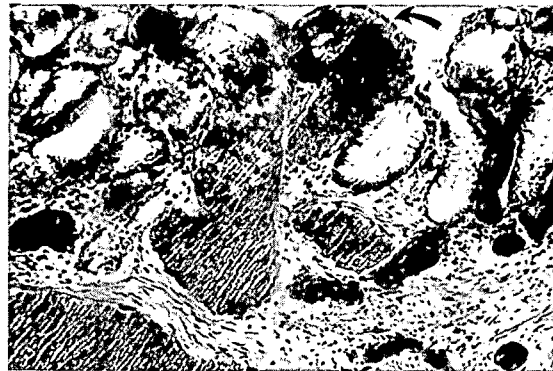
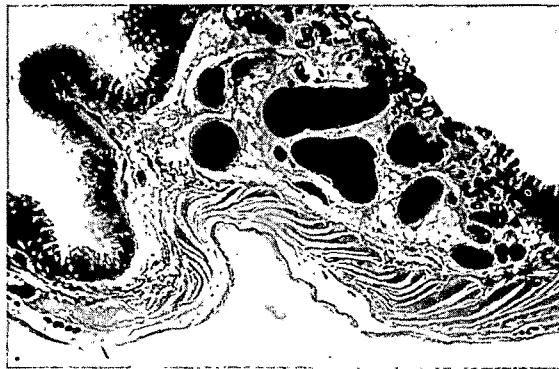
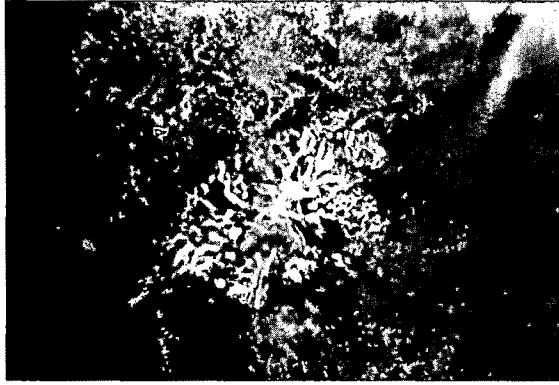


Figure 9 Angiodysplasia of the cecum. (A) Superior mesenteric arteriogram with magnification demonstrates a vascular tuft (*arrow*) and an early draining vein (*arrowheads*). (B) Late arterial phase: persistent, dense opacification of the vein. (C) A typical angiodysplasia lesion as visualized under the dissecting microscope. (D) Histological section (10x magnification) through a typical lesion shows multiple dilated submucosal vascular spaces (*black*). (E) 40x magnification of a section from the specimen in D. The dilated vascular space is separated from the colonic lumen by a single layer of cells (*arrow*).



been observed (16). During intravenous vasopressin therapy, care must be taken to use a central venous line and thus avoid dermal vasoconstriction and necrosis.

Complications can arise during the use of particulate embolic material or tissue adhesives. Embolization or occlusion of distant or neighboring circulations may result in infarction, pain, or abscess formation. This is usually a result of improper technique and can be avoided by the use of balloon occlusion techniques and/or superselective embolization.

CONCLUSION

Selective arteriography is a valuable method for the localization of sources of acute and chronic gastrointestinal bleeding. With intraarterial and intravenous infusions of vasopressin and transcatheter embolization the bleeding can be controlled in the majority of patients. In those patients in whom arteriography has failed to control the bleeding, localization of the source of bleeding facilitates the surgical procedure.

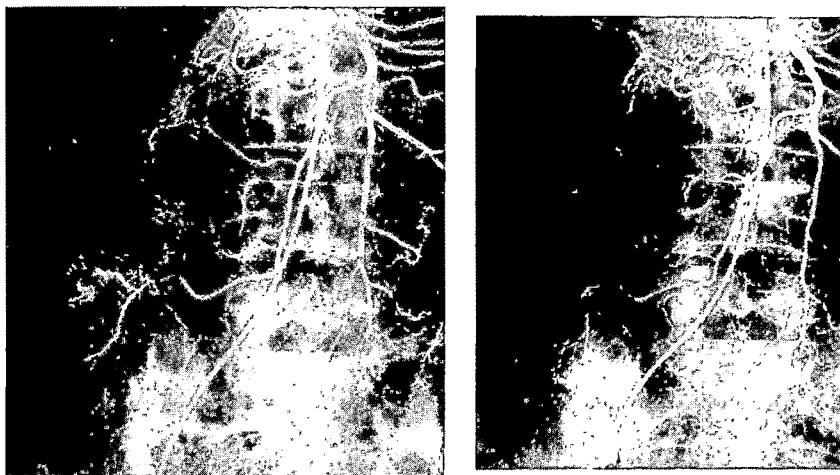


Figure 10 Postoperative cecostomy bleeding. (A) Superior mesenteric arteriogram demonstrates the bleeding site in the cecum. (B) The bleeding stopped following intraarterial infusion of vasopressin at 0.2 U/min.

Literature Cited

1. Selding, S. L. 1953. Catheter replacement of needle in percutaneous arteriography: A new technique. *Acta Radiol.* 39:368-76
2. Ödman, P. 1956. Percutaneous selective angiography of the main branches of the aorta. *Acta Radiol.* 45:1-14
3. Waltman, A. C., Courey, W. R., Athanasoulis, C. A., et al. 1973. Technique for left gastric artery catheterization. *Radiology* 109:732-34
4. Kehne, J. H., Hughes, F. A., Gomperte, N. L. 1956. The use of surgical pituitrin in the control of esophageal varix bleeding. *Surgery* 39:917-25
5. Goodman, L. S., Gilman, A. 1970. *The Pharmacological Basis of Therapeutics*. New York: MacMillan 4th ed.
6. Athanasoulis, C. A., Waltman, A. C., Novelline, R. A., et al. 1976. Intraarterial infusions of vasopressin for the control of upper gastrointestinal hemorrhage. In *Gastrointestinal Emergencies*, ed. H. R. Clearfield, V. P. Dinoso, Jr., pp. 37-47. New York: Grune & Stratton
7. Baum, S. 1974. Hematemesis and melena. Part II. Angiography. In *Gastroenterology*, ed. H. L. Bockus, pp. 827-46. Philadelphia: Saunders
8. Athanasoulis, C. A., Baum, S., Waltman, A. C., et al. 1974. Control of acute gastric mucosal hemorrhage: Intraarterial infusion of posterior pituitary extract. *N. Engl. J. Med.* 290:597-603
9. Hedberg, S., Welch, C. E. 1973. Gastrointestinal hemorrhage 1: General considerations of diagnosis and therapy. *Adv. Surgery* 7:95-148
10. Athanasoulis, C. A., Waltman, A. C., Novelline, R. A., et al. 1976. Angiography: Its contribution to emergency management of gastrointestinal hemorrhage. *Radiol. Clin. North Am.* 14: 265-80
11. Casarella, W. J., Kanter, I. E., Seaman, W. B. 1972. Right sided colonic diverticula as a cause of acute rectal hemorrhage. *N. Engl. J. Med.* 286:450-53
12. Athanasoulis, C. A., Baum, S., Rösch, J. 1975. Mesenteric arterial infusion of vasopressin for hemorrhage from colonic diverticulosis. *Am. J. Surg.* 129:212-16
13. Athanasoulis, C. A., Galdabini, J. J., Waltman, A. C., et al. 1978. Angiodysplasia of the colon: A cause of rectal bleeding. *Cardiovasc. Radiol.* 1:3-13
14. Kadir, S., Athanasoulis, C. A. 1978. Angiographic diagnosis and management of postoperative bleeding. *Crit. Rev. Med. Imaging*. In press
15. Meany, T. F. 1973. *Percutaneous Femoral Angiography: Complications and Legal Implications of Radiographic Special Procedures*, ed. T. F., Meany, R. J. Alldi, pp. 15-35. St. Louis: Mosby
16. Kadir, S., Athanasoulis, C. A. 1978. Catheter dislodgement: A cause for the failure of intraarterial vasopressin to control gastrointestinal hemorrhage. *Cardiovasc. Radiol.* 1:187-91

THE MEDICAL MANAGEMENT OF PEPTIC ULCER

◆7306

David J. C. Shearman, M.D., Ph.D. and David Hetzel, M.D.

Department of Medicine, University of Adelaide, Royal Adelaide Hospital,
Adelaide, South Australia 5000

INTRODUCTION

There is a resurgence of interest in the management of peptic ulceration because of the advent of endoscopy, which permits ulcers to be diagnosed accurately and their healing assessed, and because of the introduction of drugs that accelerate healing.

Evaluation of treatments for peptic ulcer requires carefully planned clinical trials in which patients are randomly allocated to treatment groups, preferably in a double-blind fashion. The design of the trial and interpretation of the results must take into consideration the normal course of peptic ulcer disease. Gastric and duodenal ulcers probably heal at approximately the same rate (1), and so the time taken to attain complete healing depends on the initial size of the ulcer. The duration of a trial must be sufficient to allow measurement of the healing rate by successive endoscopic measurements. Early differences between placebo and treatment groups may be obscured if endoscopic assessment is made only at the end of a trial, because ulcers tend to heal naturally regardless of the treatment. In recent clinical trials data on relapse rates in patients receiving placebo preparations suggest that relapse occurs more rapidly and more frequently than clinical assessment or radiological studies had previously estimated. Relapse rates of 65–80% within six months and 80–100% within one year have been found in these trials for both duodenal and gastric ulcers (2–4). Endoscopy has also shown that approximately one fifth of ulcers recur in the absence of symptoms (5).

Randomized clinical trials performed on duodenal ulcer patients between 1964 and 1974 have been analyzed (6) and these data, together with more recent trials, particularly of cimetidine, are discussed in this review.

ANTACID THERAPY

Antacids act by reducing gastric acidity, thereby raising gastric and duodenal pH; in addition, the activity of pepsin is almost abolished at a pH above 4. Peptic activity is also reduced by adsorption of pepsin to aluminum hydroxide (7), but it is not known whether this effect is clinically important. The duration of action of antacids is limited by gastric emptying, lasting 30 minutes in the fasting state and 2–3 hours after meals. Formulation of the antacid is of great importance. A suspension is easily solubilized and can therefore exert its neutralizing effect before it is emptied from the stomach. Tablet antacids have a poor solubility and aluminum hydroxide gel, in particular, loses most of its effect in tablet form (8).

Antacids for Symptomatic Relief

It is a frequent clinical observation that antacids alleviate ulcer pain within minutes, even when they are administered in a dosage too small to significantly affect the gastric pH. This may be a placebo response or antacids may conceivably relieve symptoms by unexplained mechanisms. For example, it has been suggested that the pain experienced by some duodenal ulcer patients may be caused by the action of acid on inflamed mucosa in the lower esophagus (9); it is possible that a small dose of antacid could relieve such symptoms. In a study in which antacid or a placebo of similar appearance and flavor was given for episodes of pain in patients with duodenal ulcer, both preparations were equally effective (10). This suggests that the expectation of pain relief is an important feature of antacid action.

The effect of large doses of antacids on the symptoms of duodenal ulcer did not differ from that of placebo over a four-week period (11, 12), yet antacid promoted healing. Again it could be concluded that antacid has an important placebo effect on pain relief.

Antacids for Healing Ulcers

Early trials (13, 14) failed to demonstrate a healing effect of antacids but these trials were either uncontrolled or used calcium carbonate, which has side effects that may prevent healing. With endoscopy and well-designed clinical trials, recent evidence suggests that antacids will heal peptic ulcers. In a double-blind trial of four weeks' duration, liquid antacid or placebo was given to 74 patients with duodenal ulcer in a dose of 30 ml one and three hours after each meal and at night (12). Of 36 patients who received antacid, 28 had complete healing of the ulcer; only 17 of 38 patients who received placebo had healed ulcers ($P < 0.005$). When the same antacid regimen was compared to cimetidine in a dose of 800–1200 mg per day, 64% of patients with duodenal ulcer receiving cimetidine and 59% of those receiving

placebo had healed, an insignificant difference (15). Antacids given in large doses have similar effects on intragastric pH to cimetidine (16), and it is not surprising therefore that both forms of treatment are effective in healing duodenal ulcers. There is also evidence to suggest that antacids promote the healing of gastric ulcers (11, 17).

Side Effects of Antacids

Diarrhea due to magnesium salts, or occasionally constipation caused by aluminum hydroxide are the main side effects that limit the use of prolonged, high dosage, antacid therapy. Calcium antacids should not be used because of their role in the milk alkali syndrome and because they produce a rebound increase in acid secretion after their neutralizing effect has subsided (18); it is possible that calcium directly induces the antral mucosa to release gastrin. Most antacids contain small amounts of sodium (Table 1), which will exacerbate fluid retention in cardiac failure, cirrhosis, and renal failure. Small amounts of magnesium and aluminum are absorbed and, in patients with renal failure, sufficient amounts may be retained to cause magnesium or aluminum toxicity. A phosphorus depletion syndrome that presents with weakness, debility, and anorexia has been reported with aluminum antacids; aluminum combines with phosphorus in the intestine and this results in reduced urinary phosphate and raised urinary calcium concentrations. There is increased resorption from bones, and osteomalacia and osteoporosis result (19). Antacids may affect the absorption of other drugs by adsorbing them or by increasing the intragastric or urinary pH (20, 21). The serum level of salicylates (22), some antibiotics, and digoxin (23) is reduced, and magnesium hydroxide will increase the absorption of bishydroxycoumarin (24). Aluminum inhibits the contraction of gastric smooth muscle, and aluminum antacids given alone will delay gastric emptying (25); this may be responsible for the delayed absorption of some drugs. Antacids do not affect the absorption of cimetidine (26).

Table 1 The volume of proprietary aluminum- and magnesium-containing antacids required to reduce gastric acidity from 50 to 25 meq/liter. This volume should be the dose administered. The amount of sodium in this volume is shown (Data quoted in Ref. 66).

Antacid	Volume (ml)	Sodium (mg) in 80 meq
Delcid®	11	26
Mylanta II®	19	31
Aludrox®	29	29
Maalox®	31	35
Creamalin®	31	19
WinGel®	36	9
Riopan®	36	5

Clinical Use of Antacids

Fordtran and colleagues (27) studied ulcer patients in whom the peak acid output in response to histamine was greater than 25 meq/liter, and found that one hour after a steak meal 80 meq of antacid were required to reduce the gastric acidity from 50 to 25 meq/liter. From these data a dose of 80 meq of antacid per hour was recommended; the volumes of different antacids necessary are listed in Table 1. An hourly regimen is poorly tolerated over a prolonged period and, if the physician decides to use antacids, the data on the healing of duodenal ulcer suggest that 80 meq of liquid antacid one and three hours after food and at bedtime is the appropriate therapy. Because of the lower acid secretion in most patients with gastric ulcer, 40 meq would seem adequate.

HISTAMINE H₂-RECEPTOR ANTAGONISTS

The histamine H₂-receptor antagonists act on histamine receptors in the stomach to reduce acid secretion. Conventional antihistamines, such as mepyramine, do not antagonize the potent stimulatory effect of histamine on gastric acid secretion, whereas they do inhibit the action of histamine on the smooth muscle of the gastrointestinal tract and bronchus. Thus two types of receptors mediating the actions of histamine can be distinguished. Those blocked by mepyramine are called H₁ receptors. Those not blocked by conventional antihistamines are called H₂ receptors and are present in the gastric mucosa and other tissues including atria and uterus (28).

Metiamide and cimetidine have a structural resemblance to histamine (Figure 1). H₁-receptor antagonists, which possess a sidechain resembling that in histamine, lack the imidazole ring.

Metiamide was shown to be very effective in inhibiting acid secretion in response to a variety of stimulants, histamine, pentagastrin, insulin, and food (29–31). It also relieved ulcer symptoms (32) and promoted healing of ulcers (33). However, bone marrow suppression with neutropenia occurred (34) and the drug was withdrawn from use (35). It was replaced by cimetidine in which a cyanoguanidine group was substituted for the thiourea portion of the metiamide side chain (Figure 1). Cimetidine has been shown to be at least as effective as metiamide in reducing acid secretion (36–39) and to date no serious side effects have been documented.

Cimetidine is absorbed efficiently from the small intestine, reaches peak levels in an hour, has a half life of about two hours and is rapidly excreted by the kidney mainly as unchanged drug and sulfoxide metabolite (37). It follows that much smaller doses will produce an adequate reduction in acid secretion in patients with renal failure (40).

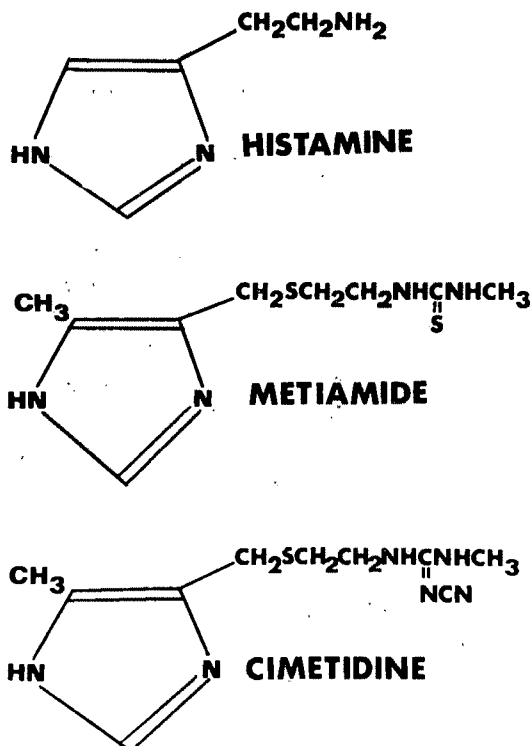


Figure 1 The structure of histamine, metiamide, and cimetidine.

Cimetidine and Duodenal Ulcer

By using endoscopy, several controlled trials from different parts of the world have shown that four to six weeks of cimetidine treatment heals the ulcer in 70–80% of patients, compared with 30–40% of patients treated with a placebo (41–45). Doses varying from 800–2000 mg daily produced similar healing rates. In all these studies cimetidine was superior to placebo in relieving symptoms and reducing the amount of antacid required for pain relief. However, the evidence from US trials is not so convincing mainly because of a higher healing rate in placebo patients. As yet there is no explanation for these differences (46) other than the possibility that peptic ulcer is a somewhat different disease in different countries. It is interesting that symptoms are relieved by cimetidine before an effect on ulcer healing can be demonstrated (41, 47): This reinforces the concept that exposure of the ulcer to acid is the cause of pain.

Although most duodenal ulcers heal rapidly on cimetidine, the cure is not permanent and patients relapse at the same rate as would be expected from the natural history of the disease (44, 48, 49).

It has now been demonstrated that continued administration of cimetidine [400 mg twice daily (3, 44, 50) or 400 mg at bedtime (4)] will prevent relapse in approximately 80% of patients over one year; it is not yet known whether this effect can be maintained indefinitely.

Cimetidine and Gastric Ulcer

In contrast to the unequivocal results of treatment of duodenal ulcer with cimetidine, therapeutic trials in chronic benign gastric ulcer have produced conflicting results. European studies, in which only small doses of antacids have been used for symptomatic relief, have found cimetidine superior to placebo when given for four to six weeks (51, 52). However a trial with small numbers of patients showed an advantage for cimetidine (80% healed at four weeks) but a high rate of healing in the placebo group (50% at four weeks) (53).

Interpretation of studies conducted in the US has been confounded by the administration of large quantities of antacid to both cimetidine and placebo groups (54) or by the failure to include a placebo-treated group (17). In the former study a trend in favor of cimetidine was evident, and in the latter 60% of patients had healed ulcers after six weeks' treatment with cimetidine (1200 mg) or with Mylanta II® (345 meq per day).

In most of these studies the influence of cimetidine on ulcer pain was also compared with placebo or antacid. In none was there any statistically significant advantage for cimetidine despite some trends in its favor. The comparison of cimetidine with other available effective treatments for gastric ulcer is important but not yet completed.

Preliminary data indicate that cimetidine (400 mg twice daily) is effective in maintaining remission of gastric ulcer (3). It must be emphasized that the healing of a gastric ulcer does not exclude malignancy (55). Repeat endoscopy with multiple biopsy and brushings is essential in the follow-up of all patients with gastric ulcer.

Cimetidine and Upper Gastrointestinal Bleeding

Cimetidine appears to be effective in arresting hemorrhage from esophageal or gastric erosions in patients suffering from acute liver failure (56) and in the prophylaxis of upper gastrointestinal bleeding after renal transplantation (57). Further work is clearly necessary, but it seems worthwhile to give cimetidine prophylactically in circumstances where severe stress is known to lead to a high incidence of hemorrhage from acute erosions.

Cimetidine and the Zollinger-Ellison Syndrome

The majority of patients with Zollinger-Ellison syndrome have their symptoms effectively controlled by treatment with H₂-receptor antagonists (58).

Symptoms of ulceration, malabsorption, diarrhea, and fluid and electrolyte loss are relieved. The basic disease process is unaffected but the alternative approach, removal of the end organ by total gastrectomy and leaving the tumor in situ, has many disadvantages and cimetidine is now the treatment of choice. However, when the tumor is confined to the duodenal wall, curative surgery may be possible.

Safety of Cimetidine

Adverse effects from cimetidine have been rare, but continuous treatment for more than a few months has not been monitored in sufficient patients for long-term safety to be assessed. An increased gastrin response to food has been described in patients with duodenal ulcer receiving cimetidine for a prolonged period (J. Hansky, personal communication), but rebound increase of acid secretion above pretreatment levels has not been shown (44, 45).

Gynecomastia has been observed in 16 patients (59), six of whom suffered from Zollinger-Ellison syndrome. Most had received more than three months' continuous treatment, but in five cases when cimetidine was continued for up to one year the breast swelling and tenderness disappeared. In those patients studied, most had normal levels of follicle-stimulating hormone, luteinizing hormone, and prolactin. However, a small number of patients with elevated serum prolactin levels has been reported (60), and acute intravenous injection of 300 mg of cimetidine may produce a transient threefold rise in serum prolactin concentration in normal men (61).

After the experience with metiamide, the possibility that granulocytopenia might occur with cimetidine was of great concern. However, the evidence suggests that bone marrow toxicity was a property of the thiourea group in metiamide (replaced by cyanoguanidine in cimetidine) rather than a product of the H_2 -receptor antagonism. Unlike metiamide, tritiated cimetidine is not incorporated into thyroid or bone marrow tissue (62) and has not caused granulocytopenia in test animals as did metiamide. The most important evidence is that three patients recovering from agranulocytosis attributed to metiamide continued to improve while receiving cimetidine (35, 59).

Transient small increases in serum creatinine occur in some 60% of patients during treatment (63, 64) but rarely raise the level above normal. The mechanism is unclear but it cannot be explained by drug interference with the laboratory test or changes in glomerular filtration; there is no evidence to suggest nephrotoxicity. Abnormal liver function tests have been found in similar numbers of placebo and cimetidine patients and clinical evidence of liver damage has been rare (59). A number of case reports associate the use of cimetidine with mental confusion, bradycardia, pan-

creatitis, muscular pain, and impaired glucose tolerance. Concomitant administration of other drugs in severely ill patients makes interpretation difficult, but an effect of cimetidine cannot be excluded.

At the dose required to produce 50% inhibition of acid secretion (approximately 2 μ mole/liter) (37) there appears to be no immediate effect on other physiological systems. The effect of prolonged therapy on H_2 receptors in other sites, such as cardiovascular, endocrine, leucocyte, and mast cell, has yet to be elucidated. Concern has been expressed about blockade of H_2 receptors on T lymphocytes, but altered immunological function has not been recognized clinically (65).

ANTICHOLINERGIC DRUGS

Anticholinergic drugs block the muscarinic effect of acetylcholine and so reduce the gastric secretion of acid. These effects on basal and stimulated secretion and their clinical use are reviewed by Piper (20) and by Littman & Pine (66). Despite their use for many years, there is no clear evidence that anticholinergic drugs relieve symptoms or heal peptic ulcers (6). The lack of evidence originates partly from the difficulty in carrying out controlled trials with drugs that cause sufficient side effects to prevent "blind" trials. For effective reduction of gastric acid secretion, it has been maintained that the dosage should be sufficient to cause side effects. This concept has been challenged by a study which showed that 15 mg of propantheline inhibited food-stimulated acid secretion as effectively as 48 mg (67). This study also demonstrated that 15 mg of propantheline augmented the inhibitory effect of 300 mg of cimetidine on food-stimulated acid secretion. A small dose of poldine methylsulfate (4 mg) will also augment the inhibition of nocturnal acid secretion by 200 or 400 mg of cimetidine (68). However, when atropine (0.6 mg) was used together with cimetidine (200 or 400 mg) no additional inhibition of intragastric acidity could be shown (69). Thus this result is at variance with the effects produced by the quaternary compounds poldine and propantheline.

Clinical Use

The use of anticholinergic drugs alone in the treatment of peptic ulceration is not recommended because their doubtful benefits do not outweigh their side effects, which include urinary retention in patients with prostatic hypertrophy, blurring of vision, the precipitation of glaucoma, gastroesophageal reflux due to reduction of tone in the lower esophageal sphincter, gastric retention in patients who have a preexisting delay in gastric emptying, and constipation. These side effects suggest avoidance of anticholinergic drugs altogether in patients over the age of 55–60. Anticholinergic drugs

should not be given as a proprietary preparation that contains antacid. Under these circumstances, the frequent use of antacid will lead to an overdosage of anticholinergic.

CARBENOXOLONE SODIUM

Carbenoxolone is a synthetic derivative of glycyrrhetic acid, a substance present in licorice root. It stimulates gastric mucus production (70), reduces the back diffusion of hydrogen ions into the mucosa (71), and increases the life span of gastric epithelial cells.

Gastric Ulcer

Since the original trial (72) there has been ample confirmation that carbenoxolone accelerates healing of gastric ulcer both in ambulant outpatients and hospital inpatients (73).

In double-blind studies lasting four weeks, carbenoxolone approximately doubled the proportion of patients with completely healed ulcers. Mean ulcer size was reduced by about 75% compared with 30–40% during placebo treatment. Most trials have allowed the use of antacids, and their use for pain relief has been significantly less in the carbenoxolone group. The main value of the drug has been to accelerate ulcer healing while allowing the patient to continue with his normal life. An effective dose with the fewest side effects is 100 mg three times a day for one week followed by 50 mg three times a day until the ulcer is healed. There is little data on the use of a maintenance dose but 50 mg daily confers no benefit (74). The use of normal doses is impractical for long-term therapy because of the hazard of adverse effects.

Duodenal Ulcer

Initial studies of duodenal ulcer did not suggest that carbenoxolone conferred any advantage over placebo. However, the early studies relied on radiology to assess healing. Trials using endoscopy suggest that delayed-release capsules containing 50 mg of carbenoxolone, given four times daily and designed to rupture in the pyloric antrum, will increase the proportion of patients with ulcers healed at six weeks and have a beneficial effect on symptoms (6, 75, 76). Although delayed release capsules were designed to deliver maximal concentration of the drug to the duodenal mucosa, it is not known whether the drug exerts its effect topically or systemically.

Adverse Effects

The usefulness of carbenoxolone is limited by its mineralocorticoid side effects—salt and water retention with edema, hypokalemia, and arterial

hypertension. These side effects are related to dose and blood levels of the drug (73, 77) and are more common in the elderly and in patients with cardiovascular, respiratory, or renal disease. Thiazide diuretics and potassium supplements will usually reverse these effects and allow continued treatment, but spironolactone antagonizes both the mineralocorticoid activity and the healing effect of the drug (78). Severe side effects including muscle weakness related to hypokalemia, congestive cardiac failure, and malignant hypertension have been reported, usually associated with prolonged use of excessive doses of the drug, often without adequate supervision of treatment (79). Clinical assessment every two weeks during treatment is necessary and the drug should be avoided in patients over the age of 65.

Deglycyrrhizinized Licorice

Deglycyrrhizinized licorice (DGL) is a licorice extract that does not cause the mineralocorticoid effects of carbenoxolone (80). There are few well-conducted trials on which to assess its value, and available reports are conflicting. Radiologically controlled trials in gastric ulcer suggest that it is superior to placebo but not as effective as carbenoxolone (81). DGL was ineffective in preventing recurrence of ulceration (82). In duodenal ulcer, symptomatic relief with DGL is only marginally better than that with placebo (6, 81).

TRIPOTASSIUM DICITRATO BISMUTHATE

Tripotassium dicitrato bismuthate (TBD) is an ammoniacal suspension of a complex colloidal bismuth salt. It is not absorbed and its mechanism of action is therefore local.

There have been a number of controlled outpatient trials employing endoscopy which show that the healing rate of both gastric and duodenal ulcers is increased (83–85). However, TBD causes black discoloration of the stool, the degree of blindness of trials may be open to doubt, and individual trials have involved only small numbers of patients. Studies with larger groups of patients over longer periods of time are required. Although no side effects have been reported, longer-term studies must assess whether any bismuth is absorbed.

GENERAL MEASURES

Dietary Therapy

There is very little evidence on the effect of diet in peptic ulcer disease, but what evidence there is suggests that dietary manipulation has no effect on

ulcer healing. The neutralizing effect of milk is small (86) and it causes secretion of acid because of its content of calcium and protein (87). Milk alone results in a lower pH in the stomach than feeding three meals a day (88). Thus milk has a transient neutralizing effect followed by a prolonged rise in acid secretion. Bland diets show no advantage over normal diets in respect to intragastric pH (89) or the natural course of the ulcer (90, 91), yet bland diets and milk remain a commonly used regimen for peptic ulcer disease in hospital practice (92).

What dietary advice should be given? Because patients with duodenal ulcer have an exaggerated secretory response to eating (93), frequent small meals should be avoided. The patient should be advised to have three normal meals each day. The excessive acid secretion in response to each meal is then controlled either by cimetidine given with meals or by antacids given one and three hours after meals. Food should not be eaten before retiring to bed since this will increase secretion at night.

Smoking

Patients with peptic ulcer are more likely to smoke than a control population and it is suspected that smoking is an etiological factor. To date there is no evidence that cessation of smoking affects the course of duodenal ulceration. The healing of gastric ulcers is aided when smoking is stopped (94) but this does not prevent their relapse (95). Nevertheless it is important to advise patients with gastric ulcer to stop smoking.

Injurious Drugs

It is now accepted that aspirin breaks the gastric mucosal barrier, thus allowing hydrogen ions to enter and damage the mucosa. Indomethacin, phenylbutazone, and ibuprofen may act in the same way. Several studies have confirmed that regular aspirin consumption is associated with chronic gastric ulcer but not with duodenal ulcer (96). Aspirin should be prohibited in patients with gastric ulcer, and because analgesic abuse is often denied it may be important to monitor the urine for aspirin.

An analysis of 6000 patients from 50 studies has failed to show an association between corticosteroid treatment and peptic ulcer (97), and there is no evidence that corticosteroids should be withheld from ulcer patients in whom they are indicated.

GUIDELINES TO THE MANAGEMENT OF PEPTIC ULCER

Recommendations on the treatment of peptic ulcer are changing rapidly as new drugs are introduced and older remedies are more thoroughly assessed.

Duodenal Ulcer

It is now apparent that a duodenal ulcer will heal if acid secretion is reduced to the same level as that brought about by surgery (98). Both cimetidine and intensive antacid therapy are capable of effecting this reduction; anticholinergic drugs are not. The cautious physician may choose intensive antacid therapy until more extensive safety data are available on cimetidine. Many physicians prefer cimetidine (200 mg three times a day with meals and 400 mg at bedtime, or 300 mg four times a day) for four to six weeks as the treatment of choice for duodenal ulcer rather than intensive antacid therapy because of intolerance to antacids given for a six week period. In one study (17) diarrhea occurred in one third of patients receiving antacids. There is also evidence that the patient does not consume the recommended dose of antacid (99). Cimetidine also has advantages over other drugs marketed for duodenal ulcer, e.g. bismuth compounds and carbenoxolone, which have not undergone such extensive trials and do not produce the rapid symptomatic relief given by cimetidine. Cimetidine is clearly the treatment of choice in peptic ulcer occurring in renal failure, for peptic ulcer in the elderly, and for those with Zollinger-Ellison syndrome.

Anticholinergic drugs should not be prescribed as the sole treatment for duodenal ulcer. There is a theoretical advantage for using them in the following situations: A small dose of anticholinergic 15–30 minutes before food will prolong the action of antacids given one to three hours after food, by slowing gastric emptying (100). In contrast to the daytime use of anticholinergic drugs, the degree of nocturnal acid suppression is related to the dose of the drug (101). Thus if intensive antacid therapy is utilized, a large dose of anticholinergic at bedtime will reduce nocturnal acid secretion, thereby controlling pain.

Cimetidine is the only drug so far demonstrated to prevent relapse in duodenal ulcer disease, but until the effects of long-term therapy are assessed further, it is recommended that the drug should be used for each exacerbation rather than as maintenance therapy. However, maintenance therapy with 400 mg at bedtime is recommended for the elderly with severe peptic ulceration; until now the alternative therapy has been gastric irradiation (102).

Gastric Ulcer

Attention should be paid to the control of smoking and analgesic ingestion. In the absence of extensive information on the use of cimetidine for gastric ulcer, intensive antacid therapy is preferred by some physicians if the patient is prepared to take this therapy diligently. Carbenoxolone is an effective therapy (not available in the US) but it should not be used in the elderly. There is no evidence to suggest that prolonged bed rest confers any additional benefit to these specific measures.

OTHER DRUGS

Many antisecretory and antiulcer agents are being studied (103). The following are of interest, but are not in general use for the treatment of peptic ulcer.

Prostaglandin Analogues

The synthetic analogues of prostaglandin E_2 , 15-methyl PGE_2 and 16,16-dimethyl PGE_2 , are orally active inhibitors of gastric acid secretion (104) and prevent the development of experimental ulcers in animals (105). Preliminary studies in man suggest they will accelerate healing of gastric (106) and duodenal ulcers (107). The usefulness of this group of drugs may be limited by their side effects, as the majority of patients in the latter study developed diarrhea, and 16,16-dimethyl PGE_2 damages the gastric mucosal barrier (108).

Gefarnate

There is some evidence based on radiological studies that gefarnate, a synthetic isoprenoid compound originally extracted from the white-headed cabbage, will promote healing of gastric ulcers (109). Direct comparative studies suggest it is not as active as carbenoxolone (110) although it has the advantage of being free of side effects. Comparison with placebo over a period of 33 months in 22 patients with chronic gastric ulcer suggested that healing occurred significantly more often in those patients receiving gefarnate (111). There is no evidence of its value in duodenal ulcer.

Urogastrone

Urogastrone is a polypeptide of 52 amino acids that is isolated from human urine. In doses of $0.25 \mu\text{g/kg}$ per hr it inhibits by 80% the acid output stimulated by pentagastrin, histamine, and insulin (112). The polypeptide has to be given parenterally and to date it has not been used in the therapy of peptic ulceration. Its amino acid composition is similar to that of epidermal growth factor (113), which raises the possibility that urogastrone might enhance tissue repair.

Secretin

Exogenous secretin administered to duodenal ulcer patients enhances neutralization of gastric acid entering the duodenum (114) because it stimulates pancreatic bicarbonate secretion (115) and inhibits gastric acid secretion (116). However, no advantage has been found for subcutaneous secretin compared to placebo in the relief of duodenal ulcer pain (117), and no increase in healing could be demonstrated in 13 patients (118). Other hormones (e.g. somatostatin) have been found to inhibit gastric secretion and

prevent ulcer formation in animals (119) but have not been fully assessed in man.

Estrogens

Although estrogens have been shown to aid duodenal ulcer healing (120, 121), they have not been shown to benefit gastric ulcer (122). The high incidence of endocrine side effects makes them unsuitable for general use.

Sodium Amylosulfate

Sodium amylosulfate (amylopectin sulfate), a synthetic mucopolysaccharide, inhibits peptic activity in man, not by acting on pepsin but by combining with tissue proteins and rendering them less digestible by pepsin (123). In one study of 19 patients it increased the rate of healing of gastric ulcer, judged radiologically, compared with placebo therapy (124). Clinical trials that assessed the effect of sodium amylosulfate on the symptoms or healing of duodenal ulcer provided conflicting results (125-127).

Drugs Acting on the Central Nervous System

Trimipramine is a tricyclic antidepressant and sedative that slightly reduces basal, pentagastrin, and histamine-stimulated gastric acid secretion (128, 129), and in one multicenter trial produced greater improvement in symptoms and more frequent endoscopic healing than placebo (130). Sulpiride, an antidepressant and antiemetic related to metoclopramide, has been claimed to be effective in the treatment of peptic ulcer (131). It inhibits gastrin release (132).

Literature Cited

1. Scheurer, U., Witzel, L., Halter, F., Keller, H. M., Huber, R., Galeazzi, R. 1977. Gastric and duodenal ulcer healing under placebo treatment. *Gastroenterology* 72:838-41
2. Blackwood, W. S., Maudgal, D. P., Northfield, T. C. 1978. Prevention by bedtime cimetidine of duodenal-ulcer relapse. *Lancet* 1:626-27
3. Bodemar, G., Walan, A. 1978. Maintenance treatment of recurrent peptic ulcer by cimetidine. *Lancet* 1:403-7
4. Gray, G. R., Smith, I. S., Mackenzie, I., Gillespie, G. 1978. Long term cimetidine in the management of severe duodenal ulcer dyspepsia. *Gastroenterology* 74:397-401
5. Hetzel, D. J., Shearman, D. J. C., Hansky, J., Korman, M. G., Hecker, R., Jackson, R., Gabb, B. W., Sheers, R. 1978. Maintenance of remission of duodenal ulcer by cimetidine: a double-blind controlled trial. *Gut* 19:A442 (Abstr.)
6. Christensen, E., Juhl, E., Tygstrup, N. 1977. Treatment of duodenal ulcer. Randomized clinical trials of a decade (1964 to 1974). *Gastroenterology* 73: 1170-78
7. Piper, D. W., Fenton, B. 1961. The adsorption of pepsin. *Am. J. Dig. Dis.* 6:134-41
8. Morrissey, J. F., Barreras, R. F. 1974. Antacid therapy. *N. Engl. J. Med.* 290:550-54
9. Earlam, R. 1975. Epigastric pain in duodenal ulcer. *Br. Med. J.* 2:558
10. Sturdevant, R. A. L., Isenberg, J. I., Secrist, D., Ansfield, J. 1977. Antacid and placebo produced similar pain relief in duodenal ulcer patients. *Gastroenterology* 72:1-5
11. Hollander, D., Harlan, J. 1973. Antacids vs placebos in peptic ulcer therapy.

- A controlled double-blind investigation. *J. Am. Med. Assoc.* 226:1181-85
12. Peterson, W. L., Sturdevant, R. A. L., Frankl, H. D., Richardson, C. T., Isenberg, J. I., Elashoff, J. D., Sones, J. Q., Gross, R. A., McCallum, R. W., Fordtran, J. S. 1977. Healing of duodenal ulcer with an antacid regimen. *N. Engl. J. Med.* 297:341-45
 13. Doll, R., Price, A. V., Pygott, F., Sanderson, P. H. 1956. Continuous intragastric milk drip in treatment of uncomplicated gastric ulcer. *Lancet* 1:70-73
 14. Baume, P. E., Hunt, J. H. 1969. Failure of potent antacid therapy to hasten healing in chronic gastric ulcers. *Aust. Ann. Med.* 18:113-16
 15. Ippoliti, A. F., Sturdevant, R. A. L., Isenberg, J. I., Binder, M., Camacho, R., Cano, R., Cooney, C., Kline, M. M., Koretz, R. L., Meyer, J. H., Samloff, I. M., Schwabe, A. D., Strom, E. A., Valenzuela, J. E., Wintroub, R. H. 1978. Cimetidine versus intensive antacid therapy for duodenal ulcer. A multicenter trial. *Gastroenterology* 74:393-95
 16. Deering, T. B., Malagelada, J.-R. 1977. Comparison of H_2 -receptor antagonist and a neutralizing antacid on postprandial acid delivery into the duodenum in patients with duodenal ulcer. *Gastroenterology* 73:11-14
 17. Englert, E., Freston, J. W., Graham, D. Y., Finkelstein, W., Kruss, D. M., Priest, R. J., Raskin, J. B., Rhodes, J. B., Rogers, A. I., Wenger, J., Wilcox, L. L., Crossley, R. J. 1978. Cimetidine, antacid, and hospitalization in the treatment of benign gastric ulcer. A multicenter double blind study. *Gastroenterology* 74:416-25
 18. Barreras, R. F. 1973. Calcium and gastric secretion. *Gastroenterology* 64:1168-84
 19. Lotz, M., Zisman, E., Bartter, F. C. 1968. Evidence for a phosphorus-depletion syndrome in man. *N. Engl. J. Med.* 278:409-15
 20. Piper, D. W. 1973. Antacid and anticholinergic drug therapy. *Clin. Gastroenterol.* 2:361-77
 21. Leading Article. 1975. How innocuous is the white medicine. *Br. Med. J.* 2:405-6
 22. Levy, G., Lampman, T., Kamath, B. L., Garrettson, L. K. 1975. Decreased serum salicylate concentrations in children with rheumatic fever treated with antacid. *N. Engl. J. Med.* 293:323-25
 23. Green, F. W., Norton, R. A., Kaplan, M. M. 1975. Pharmacology and clinical use of antacids. *Am. J. Hosp. Pharm.* 32:425-29
 24. Ambre, J. J., Fischer, L. J. 1973. Effect of coadministration of aluminium and magnesium hydroxides on absorption of anticoagulants in man. *Clin. Pharmacol. Ther.* 14:231-37
 25. Hurwitz, A., Robinson, R. G., Vats, T. S., Whittier, F. C., Herrin, W. F. 1976. Effects of antacids on gastric emptying. *Gastroenterology* 71:268-73
 26. Burland, W. L., Darkin, D. W., Mills, M. W. 1976. Effect of antacids on absorption of cimetidine. *Lancet* 2:965
 27. Fordtran, J. S., Morawski, S. G., Richardson, C. T. 1973. In vivo and in vitro evaluation of liquid antacids. *N. Engl. J. Med.* 288:923-28
 28. Chand, N., Eyre, P. 1975. Classification and biological distribution of histamine receptor sub-types. *Agents Actions* 5:277-95
 29. Wyllie, J. H., Ealding, W. D. P., Hesselbo, T., Black, J. W. 1973. Inhibition of gastric secretion in man by metiamide: A new orally active histamine H_2 -receptor antagonist. *Gut* 14:424 (Abstr.)
 30. Mainardi, M., Maxwell, V., Sturdevant, R. A. L., Isenberg, J. I. 1974. Metiamide, an H_2 -receptor blocker, as inhibitor of basal and meal-stimulated gastric acid secretion in patients with duodenal ulcer. *N. Engl. J. Med.* 291:373-76
 31. Thjodleifsson, B., Wormsley, K. G. 1974. Gastric response to metiamide. *Br. Med. J.* 2:304-6
 32. Pounder, R. E., Williams, J. G., Milton-Thompson, G. J., Misiewicz, J. J. 1975. Relief of duodenal ulcer symptoms by oral metiamide. *Br. Med. J.* 2:307-9
 33. Multicentre Trial. 1975. Treatment of duodenal ulcer by metiamide. *Lancet* 2:779-81
 34. Forrest, J. A. H., Shearman, D. J. C., Spence, R., Celestin, L. R. 1975. Neutropenia associated with metiamide. *Lancet* 1:392-93
 35. Burland, W. L., Sharpe, P. C., Colin-Jones, D. G., Turnbull, P. R. G., Bow-skill, P. 1975. Reversal of metiamide-induced agranulocytosis during treatment with cimetidine. *Lancet* 2:1085
 36. Brimblecombe, R. W., Duncan, W. A. M., Durant, G. J., Ganellin, C. R., Parsons, M. E., Black, J. W. 1975. The pharmacology of cimetidine a new histamine H_2 -receptor antagonist. *Br. J. Pharmacol.* 53:435P-36P
 37. Burland, W. L., Duncan, W. A. M., Hesselbo, T., Mills, J. G., Sharpe, P. C., Haggie, S. J., Wyllie, J. H. 1975. Phar-

- macological evaluation of cimetidine, a new histamine H_2 -receptor antagonist, in healthy man. *Br. J. Clin. Pharmacol.* 2:481-86
38. Henn, R. M., Isenberg, J. I., Maxwell, V., Sturdevant, R. A. L. 1975. Inhibition of gastric acid secretion by cimetidine in patients with duodenal ulcer. *N. Engl. J. Med.* 293:371-75
 39. Richardson, C. T., Walsh, J. H., Hicks, M. I. 1976. The effect of cimetidine, a new histamine H_2 -receptor antagonist, on meal-stimulated acid secretion, serum gastrin, and gastric emptying in patients with duodenal ulcer. *Gastroenterology* 71:19-23
 40. Canavan, J. S. F., Briggs, J. D. 1977. Cimetidine clearance in renal failure. In *Cimetidine: Proc. 2nd Int. Symp. Histamine H_2 -Receptor Antagonists*, ed. W. L. Burland, M. A. Simkins, pp. 75-80. Amsterdam-Oxford: Excerpta Medica. 392 pp.
 41. Blackwood, W. S., Pickard, R. G., Maudgal, D. P., Lawrence, D., Northfield, T. C. 1976. Cimetidine in duodenal ulcer. Controlled trial. *Lancet* 2:174-76
 42. Bodemar, G., Walan, A. 1976. Cimetidine in the treatment of active duodenal and prepyloric ulcers. *Lancet* 2:161-64
 43. Gillespie, G., Gray, G. R., Smith, I. S., MacKenzie, I., Crean, G. P. 1977. Short-term and maintenance cimetidine treatment in severe duodenal ulceration. See Ref. 40, pp. 240-47
 44. Hetzel, D. J., Hansky, J., Shearman, D. J. C., Korman, M. G., Hecker, R., Taggart, G. J., Jackson, R., Gabb, B. W. 1978. Cimetidine treatment of duodenal ulceration. Short term clinical trial and maintenance study. *Gastroenterology* 74:389-92
 45. Winship, D. H. 1978. Cimetidine in the treatment of duodenal ulcer. Review and commentary. *Gastroenterology* 74:402-6
 46. Fordtran, J. S. 1978. Placebos, antacids and cimetidine for duodenal ulcer. *N. Engl. J. Med.* 298:1081-83
 47. Hetzel, D. J., Taggart, G. J., Hansky, J., Hecker, R., Shearman, D. J. C. 1977. Cimetidine in the treatment of duodenal ulcer. *Med. J. Aust.* 1:317-19
 48. Haggie, S. J., Fermon, D. C., Wyllie, J. H. 1976. Treatment of duodenal ulcer with cimetidine. *Lancet* 1:983-84
 49. Bardhan, K. D., Blum, A., Gillespie, G., Larkworthy, W., Mekel, R., Moshal, M., Smith, P. M., Venables, C. W., Van Tongeren, J. H., Walan, A. 1977. Long term treatment with cimetidine in duodenal ulceration. *Lancet* 1:900-1
 50. Gudmand-Hoyer, E., Birger Jensen, K., Krag, E., Rask-Madsen, J., Rahbeck, I., Rune, S. J., Wulff, H. R. 1978. Prophylactic effect of cimetidine in duodenal ulcer disease. *Br. Med. J.* 1:1095-97
 51. Bader, J. P., Morin, T., Bernier, J. J., Bertrand, J., Bétourné, C., Gastard, J., Lambert, R., Ribet, A., Sarles, H., Toulet, J. 1977. Treatment of gastric ulcer by cimetidine. A multicentre trial. See Ref. 40, pp. 287-92
 52. Frost, F., Rahbek, I., Rune, S. J., Birger Jensen, K., Gudmand Hoyer, E., Krag, E., Rask-Madsen, J., Wulff, H. R., Garbøl, J., Gottlieb Jensen, K., Højlund, M., Nissen, V. R. 1977. Cimetidine in patients with gastric ulcer: a multicentre controlled trial. *Br. Med. J.* 2:795-99
 53. Ciclitira, P. J., Machell, R. J., Farthing, M. J., Dick, A. P., Hunter, J. 1977. A controlled trial of cimetidine in the treatment of gastric ulcer. See Ref. 40, pp. 283-86
 54. Dyck, W. P., Belsito, A., Flesher, B., Liebermann, T. R., Dickinson, P. B., Wood, J. M. 1978. Cimetidine and placebo in the treatment of benign gastric ulcer. A multicenter double blind study. *Gastroenterology* 74:410-15
 55. Taylor, R. H., Lovell, D., Menzies-Gow, N., La Brooy, S. J., Misiewicz, J. J. 1978. Misleading response of malignant gastric ulcers to cimetidine. *Lancet* 1:686-89
 56. MacDougall, B. R. D., Bailey, R. J., Williams, R. 1977. H_2 -receptor antagonists and antacids in the prevention of acute gastrointestinal haemorrhage in fulminant hepatic failure: Two controlled trials. *Lancet* 1:617-19
 57. Jones, R. H., Rudge, C. J., Bewick, M., Parsons, V., Weston, M. J. 1978. Cimetidine: prophylaxis against upper gastrointestinal haemorrhage after renal transplantation. *Br. Med. J.* 1:398-400
 58. McCarthy, D. M. 1978. Report on the United States experience with cimetidine in Zollinger-Ellison syndrome and other hypersecretory states. *Gastroenterology* 74:453-58
 59. Kruss, D. M., Littman, A. 1978. Safety of cimetidine. *Gastroenterology* 74:478-83
 60. Delle Fave, G. F., Tamburrano, G., De Magistris, L., Natoli, C., Santoro, M. L., Carratu, R., Torsoli, A. 1977. Gynaecomastia with cimetidine. *Lancet* 1:1319

61. Carlson, H. E., Ippoliti, A. F. 1977. Cimetidine, an H_2 -antihistamine, stimulates prolactin secretion in man. *J. Clin. Endocrinol. Metab.* 45:367-70
62. Cross, S. A. M. 1977. The localisation of histamine H_2 -receptor antagonists. In *Proc. 18th Meet. Eur. Soc. Toxicol., Edinburgh 1976*, p. 116. Amsterdam-Oxford: Excerpta Medica
63. Leslie, G. B., Walker, T. F. 1977. A toxicological profile of cimetidine. See Ref. 40, pp. 24-33
64. Sharpe, P. C., Hawkins, B. W. 1977. Efficacy and safety of cimetidine. Long-term treatment with cimetidine. See Ref. 40, pp. 358-66
65. McGregor, C. G. A., Cochran, A. J., Ogg, L. J., Gray, G. R., Smith, I. S., Gillespie, G., Forrester, J. 1977. Immunological and other laboratory studies of patients receiving short-term cimetidine therapy. *Lancet* 1:122-23
66. Littman, A., Pine, B. H. 1975. Antacids and anticholinergic drugs. *Ann. Intern. Med.* 82:544-51
67. Feldman, M., Richardson, C. T., Peterson, W. L., Walsh, J. H., Fordtran, J. S. 1977. Effect of low-dose propantheline on food-stimulated gastric acid secretion. Comparison with an "optimal effective dose" and interaction with cimetidine. *New Engl. J. Med.* 297:1427-30
68. Saunders, J. H. B., Cargill, J. M., Wormsley, K. G. 1977. Effects of cimetidine and poldine on nocturnal gastric secretion in duodenal ulcer. *Digestion* 15:452-56
69. Pounder, R. E., Hunt, R. H., Vincent, S. H., Milton-Thompson, G. J., Misiewicz, J. J. 1977. 24-hour intragastric acidity and nocturnal acid secretion in patients with duodenal ulcer during oral administration of cimetidine and atropine. *Gut* 18:85-90
70. Shillingford, J. S., Lindup, W. E., Parke, D. V., 1974. The effects of carbenoxolone on the biosynthesis of gastric glycoproteins in the rat and ferret. *Biochem. Soc. Trans.* 2:1104-7
71. Colin-Jones, D. G., Taylor, J. V. 1973. Effect of carbenoxolone sodium upon ionic movement across the gastric mucosa in patients with gastric ulcers. *Gut* 14:423
72. Doll, R., Hill, I. D., Hutton, C., Underwood, D. J. 1962. Clinical trial of a triterpenoid liquorice compound in gastric and duodenal ulcer. *Lancet* 2:793-96
73. Pinder, R. M., Brogden, R. N., Sawyer, P. R., Speight, T. M., Spencer, R., Avery, G. S. 1976. Carbenoxolone: A review of its pharmacological properties and therapeutic efficacy in peptic ulcer disease. *Drugs* 11:245-307
74. Langman, M. J. S. 1968. The medical treatment of gastric and duodenal ulcer. *Postgrad. Med. J.* 44:603-7
75. Davies, W. A., Reed, P. I. 1977. Controlled trial of Duogastrone in duodenal ulcer. *Gut* 18:78-83
76. Nagy, G. S. 1978. Evaluation of carbenoxolone sodium in the treatment of duodenal ulcer. *Gastroenterology* 74: 7-10
77. Baron, J. H., Gribble, R. J. N., Rhodes, C., Wright, P. A. 1978. Serum carbenoxolone in patients with gastric and duodenal ulcer. Absorption, efficacy and side effects. *Gut* 19:330-35
78. Doll, R., Langman, M. J. S., Shawdon, H. H. 1968. Treatment of gastric ulcer with carbenoxolone: antagonistic effect of spironolactone. *Gut* 9:42-45
79. Davies, G. J., Rhodes, J., Calcraft, B. J. 1974. Complications of carbenoxolone therapy. *Br. Med. J.* 3:400-2
80. Cooke, W. M., Baron, J. H. 1971. Metabolic studies of deglycyrrhizinised liquorice in two patients with gastric ulcer. *Digestion* 4:264-68
81. Brogden, R. N., Speight, T. M., Avery, G. S. 1974. Deglycyrrhizinised liquorice: A report of its pharmacological properties and therapeutic efficacy in peptic ulcer. *Drugs* 8:330-39
82. Hollanders, D., Green, G., Woolf, I. L., Boyes, B. E., Wilson, R. Y., Cowley, D. J., Dymock, I. W. 1978. Prophylaxis with deglycyrrhizinised liquorice in patients with healed gastric ulcer. *Br. Med. J.* 1:148
83. Brogden, R. N., Pinder, R. M., Sawyer, P. R., Speight, T. M., Avery, G. S. 1976. Tri-potassium di-citrate bismuthate: A report of its pharmacological properties and therapeutic efficacy in peptic ulcer. *Drugs* 12:401-11
84. Coughlin, G. P., Kupa, A., Alp, M. H. 1977. The effect of tri-potassium di-citrate bismuthate (De-Nol) on the healing of chronic duodenal ulcers. *Med. J. Aust.* 1:294-98
85. Lee, S. P., Nicholson, G. I. 1977. Increased healing of gastric and duodenal ulcers in a controlled trial using tri-potassium dicitrate-bismuthate. *Med. J. Aust.* 1:808-12
86. Piper, D. W. 1969. Milk in treatment of gastric disease. *Am. J. Clin. Nutr.* 22:191-95
87. Ippoliti, A. F., Maxwell, V., Isenberg, J. I. 1976. The effect of various forms of milk on gastric-acid secretion. Studies

- in patients with duodenal ulcer and normal subjects. *Ann. Intern. Med.* 84:286-89
88. Bingle, J. P., Lennard-Jones, J. E. 1960. Some factors in the assessment of gastric antisecretory drugs by a sampling technique. *Gut* 1:337-44
 89. Lennard-Jones, J. E., Babouris, N. 1965. Effect of different foods on the acidity of the gastric contents in patients with duodenal ulcer. I. A comparison between two 'therapeutic' diets and freely-chosen meals. *Gut* 6:113-17
 90. Lawrence, J. S. 1952. Dietetic and other methods in the treatment of peptic ulcer. *Lancet* 1:482-85
 91. Doll, R., Friedlander, P., Pygott, F. 1956. Dietetic treatment of peptic ulcer. *Lancet* 1:5-9
 92. Welsh, J. D. 1977. Diet therapy of peptic ulcer disease. *Gastroenterology* 72:740-45
 93. Fordtran, J. S., Walsh, J. H. 1973. Gastric acid secretion rate and buffer content of the stomach after eating. Results in normal subjects and in patients with duodenal ulcer. *J. Clin. Invest.* 52: 645-57
 94. Doll, R., Jones, F. A., Pygott, F. 1958. Effect of smoking on the production and maintenance of gastric and duodenal ulcers. *Lancet* 1:657-62
 95. Piper, D. W., Greig, M., Coupland, G. A. E., Hobbin, E., Shinnors, J. 1975. Factors relevant to the prognosis of chronic gastric ulcer. *Gut* 16:714-18
 96. Duggan, J. M. 1976. Aspirin in chronic gastric ulcer: an Australian experience. *Gut* 17:378-84
 97. Conn, H. O., Blitzer, B. L. 1976. Nonassociation of adrenocorticosteroid therapy and peptic ulcer. *N. Engl. J. Med.* 294:473-79
 98. Pounder, R. E., Williams, J. G., Milton-Thompson, G. J., Misiewicz, J. J. 1975. 24-hr control of intragastric acidity by cimetidine in duodenal ulcer patients. *Lancet* 2:1069-72
 99. Roth, H. P., Berger, D. G. 1960. Studies on patient cooperation in ulcer treatment. I. Observation of actual as compared to prescribed antacid intake on a hospital ward. *Gastroenterology* 38: 630-33
 100. Fordtran, J. S. 1973. Reduction of acidity by diet, antacids and anticholinergic agents. In *Gastrointestinal Disease*, ed. M. H. Sleisenger, J. S. Fordtran, pp. 718-42. Philadelphia: Saunders. 1659 pp.
 101. Freston, J. W., Forbes, J. A. 1977. A double-blind evaluation of the nocturnal antisecretory effects of anisotropic methyl-bromide in man. Base response and duration of action studies. *J. Clin. Pharmacol.* 17:29-36
 102. Findlay, J. M., Newaishy, G. A., Siracus, W., McManus, J. P. A. 1974. Role of gastric irradiation in management of peptic ulceration and oesophagitis. *Br. Med. J.* 3:769-71
 103. Bass, P. 1974. Gastric antisecretory and antiulcer agents. *Adv. Drug Res.* 8:205-329
 104. Karim, S. M. M., Carter, D. C., Bhana, D., Ganesan, P. A. 1973. Effect of orally administered prostaglandin E_2 and its 15-methyl analogues on gastric secretion. *Br. Med. J.* 1:143-46
 105. Robert, A., Schultz, J. R., Nezamis, J. E., Lancaster, C. 1976. Gastric antisecretory and antiulcer properties of PGE_2 , 15-methyl PGE_2 and 16,16-dimethyl PGE_2 . Intravenous, oral and intrajejunal administration. *Gastroenterology* 70:359-70
 106. Fung, W. P., Karim, S. M. M. 1976. Effect of 15(R)15 methyl prostaglandin E_2 on the healing of gastric ulcers. A double-blind endoscopic study. *Med. J. Aust.* 2:127-28
 107. Gibinski, K., Rybicka, J., Mikos, E., Nowak, A. 1977. Double-blind clinical trial on gastroduodenal ulcer healing with prostaglandin E_2 analogues. *Gut* 18:636-39
 108. O'Brien, P. E., Carter, D. C. 1975. Effect of gastric secretory inhibitors on the mucosal barrier. *Gut* 16:437-42
 109. Newcomb, P. B., Stone, W. D., Richardson, P. C., Smith, P. M., Gent, A., Donovan, B. 1976. A double blind trial of gefarnate and placebo in the treatment of gastric ulcer. *Practitioner* 217:435-38
 110. Langman, M. J. S., Knapp, D. R., Wakley, E. J. 1973. Treatment of chronic gastric ulcer with carbenoxolone and gefarnate: a comparative trial. *Br. Med. J.* 3:84-86
 111. Truelove, S. C., Rocca, M. 1976. Treatment of chronic gastric ulcer with gefarnate: a long-term controlled therapeutic trial. *Curr. Med. Res. Opin.* 4:218-22
 112. Elder, J. B., Ganguli, P. C., Gillespie, I. E., Gerring, E. L., Gregory, H. 1975. Effect of urogastone on gastric secretion and plasma gastrin levels in normal subjects. *Gut* 16:887-93
 113. Bower, J. M., Camble, R., Gregory, H., Gerring, E. L., Willshire, I. R. 1975. The inhibition of gastric acid secretion by epidermal growth factor. *Experientia* 31:825-26

114. Andersson, S., Grossman, M. I. 1966. Effects of histalog and secretin on gastroduodenal profile of pH, potential difference and pressure in man. *Gastroenterology* 51:10-17
115. Brooks, A. M., Grossman, M. I. 1970. Effect of secretin on fasting and postprandial pH of the duodenal bulb in dogs. *Gastroenterology* 59:90-94
116. Berstad, A., Petersen, H. 1970. Dose-response relationship of the effect of secretin on acid and pepsin secretion in man. *Scand. J. Gastroenterol.* 5:647-54
117. Høj, L., Holst, J. J., Rune, S. J. 1973. A trial of exogenous secretin in the treatment of duodenal ulcer pain. *Scand. J. Gastroenterol.* 8:279-81
118. Henni, R. M., Selcon, S., Sturdevant, R. A. L., Isenberg, J. I., Grossman, M. I. 1976. Experience with synthetic secretin in the treatment of duodenal ulcer. *Am. J. Dig. Dis.* 21:921-25
119. Konturek, A. J., Radecki, T., Pucher, A., Coy, D. H., Schally, A. V. 1977. Effect of somatostatin on gastro-intestinal secretions and peptic ulcer production in cats. *Scand. J. Gastroenterol.* 12:379-83
120. Truelove, S. C. 1960. Stilboestrol, phenobarbitone and diet in chronic duodenal ulcer. *Br. Med. J.* 2:559-66
121. Siurala, M., Varis, K., Gorbatow, O., Kause, P., Laulajainen, M., Nyberg, L. O. 1969. The effect of estriol succinate upon healing of duodenal ulcer: a controlled study. *Scand. J. Gastroenterol.* 4:469-71
122. Doll, R., Langman, M. J. S., Shawdon, H. H. 1968. Treatment of gastric ulcer with oestrogens. *Gut* 9:46-47
123. Berstad, A. 1974. The effect of amylopectin sulfate (SN-263) on the peptic activity of human gastric juice. *Acta Hepato-Gastroenterol.* 21:212-20
124. Zimmon, D. S., Miller, G., Cox, G., Tesler, M. A. 1969. Specific inhibition of gastric pepsin in the treatment of gastric ulcer. *Gastroenterology* 56:19-23
125. Sun, D. C. H. 1973. A controlled trial of antacid, propantheline and amylopectin sulfate in duodenal ulcer. *Am. J. Gastroenterol.* 60:449-58
126. Landecker, K. D., McCallum, E. M., Fevre, D. I., Green, P. H. R., Kasumi, A., Piper, D. W. 1976. Effect of sodium amylosulfate (depepsen) on the healing of duodenal ulcer. *Gastroenterology* 71:723-25
127. Littman, A. 1972. Does amylopectin sulfate work in duodenal ulcer or not? (Editorial). *Gastroenterology* 62:141-42
128. Myren, J., Berstad, A. 1975. The early effect of trimipramine (Surmontil) on gastric secretion in man. *Scand. J. Gastroenterol.* 10:817-19
129. Bohmann, T., Schrumpf, E., Myren, J. 1977. The effect of trimipramine (Surmontil®) on gastric secretion and serum gastrin release in healthy young students. *Scand. J. Gastroenterol.* Suppl. 43, pp. 7-17
130. Wetterhus, S., Aubert, E., Berg, C. E., Bjerkeset, T., Halvorsen, L., Hovdenak, N., Myren, J., Roland, M., Sigstad, H., Guldahl, M. 1977. The effect of trimipramine (Surmontil®) on symptoms and healing of peptic ulcer. *Scand. J. Gastroenterol.* Suppl. 43, pp. 33-38
131. Bank, S., Marks, I. N. 1973. Evaluation of new drugs for peptic ulcer. *Clin. Gastroenterol.* 2:379-95
132. McGuigan, J. E. 1978. Sulpiride: an inhibitor of gastrin release. *Gastroenterology* 74:324-26

VALIDITY OF THERAPY FOR MILD HYPERTENSION

◆7307

Edward D. Freis, M.D.

Veterans Administration Hospital and Department of Medicine,
Georgetown University School of Medicine, Washington, DC 20422

Hypertension affects at least one in ten adults in nearly every country in the world. It ranks third only to atherosclerosis and cancer as a cause of death. It is also a leading risk factor in atherosclerotic complications such as myocardial infarction, sudden death, and atherothrombotic stroke. Hypertension, therefore, is a double-edged sword. It produces complications specifically related to hypertension per se, such as hemorrhagic stroke, congestive heart failure, and renal failure, and, in addition, it aggravates and accelerates atherosclerosis.

RATIONALE OF TREATMENT

The ultimate cause or causes of essential hypertension are still unknown. Except for rare secondary forms the goal in treating hypertension is not to cure the disorder but to control the blood pressure and, thereby, prevent hypertensive complications. This approach is based on the theory, which we can now regard as established from both animal experiments and clinical trials, that the major cardiovascular complications associated with hypertension are the result of damage produced directly by the elevated blood pressure. It has not been demonstrated, however, that reducing the elevated blood pressure will retard the development of the atherosclerotic complications often associated with hypertension.

The cardiovascular changes that characterize hypertension appear to be secondary to elevated blood pressure (1). Fibrinoid necrosis and hyalinosis of the arterioles probably represent adaptations to the increased blood pressure; cerebral microaneurysms, a frequent source of cerebral hemor-

rhage, may result from yield stresses produced by the hypertension; and left ventricular hypertrophy and dilatation represent adjustment to an increased afterload. The predisposition to atherosclerosis may be due to injury to the arterial intima caused by the hypertension (2). The rationale of antihypertensive treatment, therefore, is to reduce the blood pressure to a level where these damaging effects will not occur.

EVIDENCE IN ANIMALS

The complications of hypertension can be completely prevented by antihypertensive drug treatment in the spontaneously hypertensive rat (3). The Kyoto strain of spontaneously hypertensive rats develop a progressive elevation of blood pressure with increasing age; after one year they begin to die from cardiovascular complications associated with the hypertension. If antihypertensive agents are added to the drinking water their blood pressures remain at low levels and they develop none of the cardiovascular complications associated with hypertension. Furthermore, their life span increases by more than half that of the untreated animals and becomes the same as that of a normal rat (4). The treated animals tolerated the drugs very well; they gain weight normally, bear normal litters, and appear as active and healthy as normotensive untreated rats.

The cardiovascular complications occurring in spontaneously hypertensive rats are in many respects similar to those found in hypertensive patients and include hemorrhagic stroke, congestive heart failure, and renal damage. The rat lesions differ from those in man, however, in one important respect. The rat is resistant to atherosclerosis and seldom exhibits such lesions even when plasma cholesterol levels are increased by diet. The complete protection from cardiovascular lesions provided by treatment in the rat, therefore, applies only to hypertensive complications and not to atherosclerotic lesions, which rarely occur in the spontaneously hypertensive rat.

EARLY CLINICAL TRIALS

The evidence in man is based largely on controlled clinical trials. The first trial was carried out by Hamilton and co-workers (5) in 61 patients of both sexes with severe essential hypertension. Patients were assigned alternately to either active drugs or to no treatment. Over a follow-up period of two to six years, there were significantly fewer complications in the treated patients than in the untreated group. Almost all of the treated patients who developed complications failed to achieve a reduction of diastolic blood pressure below 110 mm Hg. Most of the complications were of the hypertensive rather than the atherosclerotic type.

A controlled trial carried out by Wolff & Lindeman (6) in 87 patients with moderate to severe hypertension indicated that over an average follow-up period of two years the incidence of morbid events in the treated patients was one third that observed in the placebo group.

Carter (7) carried out a prospective randomized trial in 97 hypertensive stroke survivors. Over a period of three to five years 46% of the control patients died as compared to 26% of the treated group. Nonfatal strokes occurred in 23% of the control group as compared to 14% of the treated patients. However, in a well-controlled trial in patients with mild hypertension who had survived a single stroke, Hoobler and his associates (8) found no significant protection against subsequent strokes by treatment.

THE VETERANS ADMINISTRATION TRIAL

The largest scale controlled trial was that carried out by the Veterans Administration Cooperative Study Group (9-11). This randomized, double-blind study included 523 male hypertensive patients, average age 49 years, whose initial diastolic blood pressure ranged between 90 and 129 mm Hg (average of two clinic visits immediately preceding randomization). Many of the patients showed evidence of end-organ disease and all maintained a diastolic blood pressure of 90 mm Hg or higher during a week's hospitalization. Patients were randomly assigned, double-blind either to a combination of hydrochlorothiazide, reserpine, and hydralazine or to placebo of these drugs.

The trial was terminated early in the subgroup of 143 patients with initial diastolic blood pressure averaging between 115 and 129 mm Hg (9). Of the 70 patients in the control group 27 developed major cardiovascular complications over an average follow-up period of only 20 months. Most of these complications were of the hypertensive rather than the atherosclerotic type. In the treated group of 73 patients there was only one major cardiovascular event, which was an atherothrombotic stroke. Two others were removed from the study because of side effects of the antihypertensive drugs.

There remained in the trial 380 patients with mild to moderate elevations of blood pressure prior to randomization averaging between 90 and 114 mm Hg (10). These patients were followed for an average period of 3.3 years and some were followed for more than five years. Nineteen of the control patients died of cardiovascular complications as compared to eight in the treated group, a ratio of more than two to one in favor of treatment. All of the deaths occurring in the treated group of patients were related to atherosclerotic complications whereas the deaths in the control group were due to either hypertensive or atherosclerotic events. Since many of the

patients had hypertension of long duration, it would be expected that they already had extensive atherosclerosis at the time of entry.

The incidence of all morbid events both nonfatal as well as fatal was assessed by the life-table method of analysis (Figure 1). This indicated that over a five-year period the risk of developing a major cardiovascular complication was reduced from 55% in the control group to 18% in the treated patients, a ratio of approximately three to one in favor of treatment. In addition to these morbid events there were 20 patients, all in the control group, who exhibited progression of their hypertension to above 120 mm Hg diastolic blood pressure and who were removed from the trial before they developed a major complication.

The effectiveness of treatment was definitely related to the initial height of the blood pressure (10, 11). The striking effect of treatment in the group

ALL MORBID EVENTS

Estimated cumulative incidence of morbidity over a five-year period as calculated by life-table method.

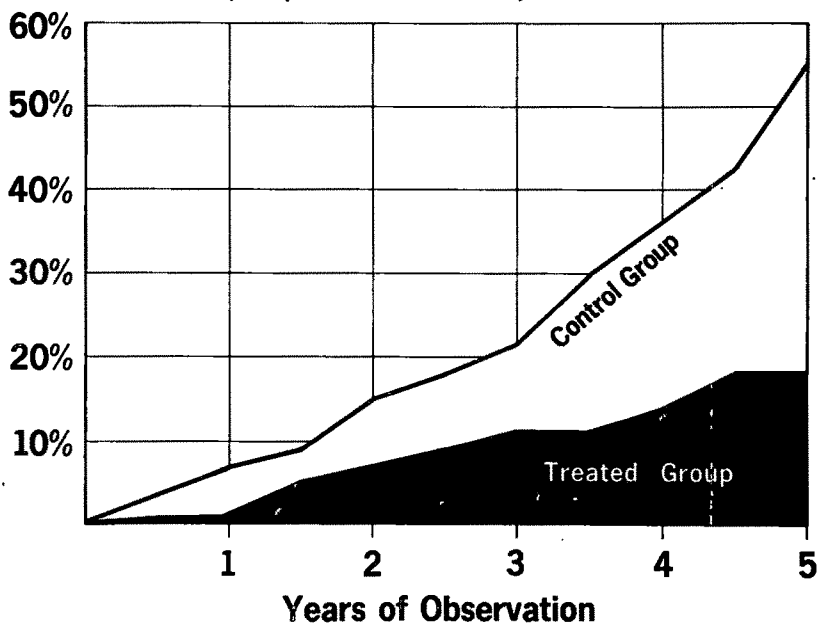


Figure 1 Graph of estimated five-year morbidity in the control vs the treated group of 380 male patients with initial diastolic blood pressure in the range of 90–114 mm Hg inclusive. (Data from Ref. 10).

with severe hypertension has already been described. The patients with mild to moderate hypertension were subdivided into two groups, those with initial diastolic levels of 90–104 mm Hg and the group with diastolic blood pressures of 105–114 mm Hg at entrance into the study. Treatment was considerably more effective in the latter than in the former or mild group. In the patients with moderate hypertension cardiovascular complications were four times higher in the control than in the treated patients. But, in the mild group the difference was less than two to one in favor of treatment. Because of this minor difference in incidence of complications and also because of the relatively small size of the mild subsample, the effectiveness of treatment for mild hypertension was left in doubt. It is worth noting, however, that progression to a more severe stage of hypertension and the development of left ventricular hypertrophy were not included as assessable morbid events. If they had been included the results would have indicated a greater effectiveness of treatment.

US PUBLIC HEALTH SERVICE HOSPITALS TRIAL

There are, as yet, no other controlled trials on mild hypertension in which the data have been published in full. The results of the study in the United States Public Health Service Hospitals conducted by Dr. W. McFate Smith have been published only in preliminary form (12). Their patients numbered 389 and were followed for as long as seven years. In contrast to the Veterans Administration study, none of their patients exhibited evidence of end-organ disease at the time of entry into the study. Furthermore, they tended to be younger with an average age of 44 years as opposed to 49 years in the Veterans trial. The trial also included females as well as male patients. While they admitted patients with diastolic levels ranging between 90 and 115 mm Hg, the majority of the patients had rather mild hypertension as indicated by the average admission blood pressure of 148/99 mm Hg.

Over the seven-year period of follow-up, hypertensive complications occurred more than twice as frequently in the control as in the treated group. Most of the preventable complications appeared to be related to electrocardiographic manifestations of left ventricular hypertrophy. The incidence of atherosclerotic events, however, including those related to coronary artery disease, was essentially the same in the treated and control patients.

CORONARY ARTERY DISEASE

Hypertension even of mild degree aggravates and accelerates atherosclerosis (13). Atherosclerosis appears to be a response to injury to the arterial

endothelium (2) that in many instances could be the result of the hypertension. Hypertension also causes structural changes in the arterial walls such as disruption of the elastic lamellae and proliferation of fibrous and collagenous connective tissue with resulting loss of distensibility of the arterial wall. The distention of arteries secondary to hypertension also may stretch the endothelium to make it more permeable to lipids. The question of reversibility of these changes with continued reduction of blood pressure has not yet been settled.

In the Veterans Administration trial, treatment was highly effective in preventing the complications that are specifically related to the hypertension itself as opposed to atherosclerotic complications. Hypertensive complications such as cerebral hemorrhage, renal functional deficits, congestive heart failure, and dissecting aortic aneurysm occurred exclusively in the control group. On the other hand, atherosclerotic events such as myocardial infarction, atrial fibrillation, and heart blocks appeared with approximately the same frequency in the control and treated patients. Smith found essentially the same differential effect of treatment on hypertensive as compared to atherosclerotic complications in patients with mild hypertension.

It might be argued that many of the patients entering the Veterans study were not representative of the general hypertension population because of age, hypertension of long standing, and evident target-organ disease. Therefore, many of them may have had extensive sclerosis of their coronary arteries at the time of entry and the risk of myocardial infarction was great regardless of the level of blood pressure. However, as noted above, Smith's study in patients with uncomplicated mild hypertension provides similar results; the incidence of complications due to coronary artery disease was not significantly different in the treated and control groups (12).

Neither the Veterans Administration study nor the Public Health Service Hospital trial were specifically designed to answer the question of the effect of treatment in preventing myocardial infarction. Larger scale studies currently underway in the United States and other countries should provide more definitive evidence. For the present, it seems best to keep an open mind on the subject, although the prospects for achieving significant benefit with blood pressure reduction alone do not appear promising.

Myocardial infarction is far and away the leading cause of disability and death in patients with mild hypertension. Such patients rarely develop the hypertensive complications that are so spectacularly benefited by treatment. The usual cause of death in untreated mild hypertension is either myocardial infarction or sudden death; if the patients survive to old age they may also die of atherothrombotic stroke.

DECISION TO TREAT MILD HYPERTENSION

Mild hypertension is herein defined as a diastolic blood pressure in the range of 90–104 mm Hg inclusive (fifth phase of Korotkoff sounds) over three or more successive office or clinic visits. The reasons for stipulating at least three visits is that the readings on the initial visit may be misleadingly high and also that blood pressure characteristically fluctuates over a wide range.

It must be admitted at the outset that our knowledge of the value of treatment in mild hypertension is fragmentary. Yet, it is far and away the most frequently encountered form of hypertension in clinical practice. According to the National Health Survey, the prevalence of mild hypertension is approximately five times higher than all other forms combined (14). The problem is ever with us and we must, therefore, make as prudent decisions as possible. To decide to treat all individuals with diastolic levels of 90 mm Hg or above would involve about 20 million patients in the United States (estimated number with such average levels over three visits), who would be receiving antihypertensive agents on a life-long basis. The cost and risks of such an undertaking are staggering. Before doing so we must be certain that the therapeutic program is justified. If drug therapy is of only questionable value we should ask whether its benefits justify the expense, inconvenience, and possible toxic side effects associated with treatment. The common thread that ties together the results of the various therapeutic trials is that treatment is highly effective against hypertensive complications but is only questionably effective against atherosclerotic complications.

Treatment, therefore, is definitely indicated whenever the risk of hypertensive complications is present. Also, the several controlled trials indicated that approximately 1–2% per year of mild hypertensives progress to a more severe stage. In the Veterans Administration trial this progression occurred mostly in younger patients. Also, the evidence is clear that electrocardiographic changes indicative of left ventricular hypertrophy are favorably influenced by treatment.

Patients with mild hypertension who are at greatest risk of developing hypertensive complications or progression of hypertension are those with the following characteristics: (*a*) presence of target-organ disease such as LVH, Group II fundi, etc; (*b*) age below 45 years; (*c*) black race; (*d*) males; (*e*) diastolic pressure above 95 mm Hg on each of three pretreatment visits; and (*f*) family history of severe hypertensive disease.

The more of these hypertensive risk factors that are present, the greater is the indication for treatment. Also, the more closely the average diastolic blood pressure approaches 105 mm Hg, the more likely it is that treatment

will be undertaken. Both the height of the blood pressure and the number of other risk factors present are essential components in the therapeutic formula.

The remaining individuals with mild hypertension are mostly those above the age of 45 who are without symptoms. This represents the largest group of the hypertensive population and their chief risk is the development of atherosclerotic complications. Although there is no proven effective preventive treatment for these patients it would seem prudent to utilize whatever knowledge we have short of exposing them to drugs. Such "hygienic" measures would include low-sodium and low-saturated-fat diets, regular exercise, weight reduction of the obese, and discontinuation of cigarette smoking.

A more aggressive approach will, of course, be justified if future therapeutic trials produce evidence that antihypertensive drug treatment protects against the complications of coronary artery disease. Several recent controlled trials have indicated that beta-adrenergic-blocking drugs reduce the risk of myocardial infarction or sudden death (15, 16). While these studies need confirmation, they suggest that reducing the adrenergic drive to the heart may have protective effects independent of, or in addition to, simple reduction of blood pressure.

SUMMARY

Hypertensive patients are candidates for drug therapy when they are at risk of developing hypertensive complications. There is at present no evidence that drug therapy is effective when the risk is limited only to atherosclerotic complications. However, the question must be left open because of insufficient data.

Potential benefits of drug therapy must be weighed against potential risks, expense, and inconvenience. Patients at high risk of hypertensive complications such as young black males probably should be treated even if the hypertension is mild. On the other hand, in patients with mild hypertension who are not at such risk it would seem prudent to utilize dietary and hygienic measures that may reduce atherosclerotic changes and that are not harmful in themselves. Whatever decision is made regarding drug treatment the patients must be followed at regular intervals because of the tendency for some to progress to a more severe stage.

Literature Cited

1. Pickering, G. W. 1968. *High Blood Pressure*, pp. 310-13, 351-52. New York: Grune & Stratton. 717 pp. 2nd ed.
2. Ross, R., Glomset, J. A. 1976. The pathogenesis of atherosclerosis. *N. Engl. J. Med.* 295:369-77
3. Freis, E. D., Ragan, D., Pillsbury, H., Mathews, M. 1972. Alteration in the course of hypertension in the spontaneously hypertensive rat. *Circ. Res.* 41:1-7
4. Freis, E. D., Ragan, D. 1976. Treatment and longevity in spontaneously hypertensive rats (SHR). *Clin. Exp. Pharmacol. Physiol.* Suppl. 3, pp. 75-78
5. Hamilton, M., Thompson, E. N., Wisniewski, T. K. M. 1964. The role of blood pressure control in preventing complications in hypertension. *Lancet* 1:235-38
6. Wolff, F. W., Lindeman, R. D. 1966. Effects of treatment in hypertension: Results of a controlled study. *J. Chronic Dis.* 19:227-40
7. Carter, A. B. 1970. Hypotensive therapy in stroke survivors. *Lancet* 1: 485-89
8. Hypertension-Stroke Cooperative Study Group. 1973. Stroke recurrence after antihypertensive therapy. *Circulation* 48: Suppl. IV-84 (Abstr.)
9. Veterans Administration Cooperative Study Group on Antihypertensive Agents. 1967. Effects of treatment on morbidity: Results in patients with diastolic blood pressures averaging 115 through 129 mm Hg. *J. Am. Med. Assoc.* 202:1028-34
10. Ibid 1970. II. Results in patients with diastolic blood pressure averaging 90 through 114 mm Hg. *J. Am. Med. Assoc.* 213:1143-52
11. Ibid 1972. III. Influence of age, diastolic pressure and prior cardiovascular disease; further analysis of side effects. *Circulation* 45:991-1004
12. US Public Health Service Hospitals Cooperative Study Group. Smith, W. M. (Chairman). 1977. Treatment of mild hypertension. Results of a ten-year intervention trial. *Circ. Res.* 40: Suppl. 1, pp. 98-105
13. Freis, E. D. 1969. Hypertension and atherosclerosis. *Am. J. Med.* 46:735-40
14. National Health Survey. 1966. Hypertension and hypertensive heart disease in adults. 1960-62, US Dept. Health, Educat. Welfare, Vital Health Stat. Ser. II, No. 13. Washington, DC: GPO
15. Wilhelmsson, C., Vedin, J. A., Wilhelmssen, L., Tibblin, G., Werkö, L. 1974. Reduction of sudden deaths after myocardial infarction by treatment with alprenolol. Preliminary results. *Lancet* 20:1157-60
16. Ahlmark, G., Saetre, H., Korsgren, M. 1974. Reduction of sudden deaths after myocardial infarction. Letter to the editor. *Lancet* 2:1563

SICK SINUS SYNDROME

♦7308

J. Thomas Bigger, Jr., M.D. and James A. Reiffel, M.D.¹

Department of Medicine, Columbia University, New York, NY 10032

INTRODUCTION

Between 1900 and 1910, Keith & Flack (1) discovered the sinoatrial (S-A) node. Flack cooled mammalian hearts and found that the heart slowed only when the cooling was applied to the region of the S-A node (2); he concluded that this structure was the pacemaker of the heart. Between 1910 and 1920, Eyster & Meek (3-5), Lewis (6, 7), and Wybauw (8) studied the sequence of activation of the heart by recording directly from its surface. These workers found that the initial electrical activation of the heart ("the area of primary negativity") occurred in the vicinity of the S-A node, a finding which suggests that the heart beat originates in the S-A node. In 1952, Trautwein & Zink recorded the transmembrane voltage time course of pacemaker (P) cells in the frog sinus venosus (9), and in 1955 West made intracellular recordings from the P cells in the rabbit S-A node (10). In accord with previous hypotheses, these workers demonstrated directly the pacemaker activity of sinus node cells.

Disorders of human S-A node pacemaker function were reported by Wenckebach (11) and Laslett before 1910 (12). Sinoatrial block in man was reported by Eyster & Evans in 1915 (13) and by Levine in 1916 (14). However, human S-A node disorders have only been widely recognized during the past 10 years. In recent years, the term sick sinus syndrome has been used to describe an array of clinical disorders of S-A node function in man (15-17). Other terms used to describe these disorders include sluggish sinus node syndrome (18), lazy sinus node (19), inadequate sinus mechanism (20), and sinoatrial syncope (21). When supraventricular tachyarrhythmias recur in patients with sinus bradycardia and/or brady-

¹Supported in part by grants (HL12738, HL70204, and HL05864) from the National Institutes of Health and by grants-in-aid from the Chernow Foundation and from the Winthrop Foundation.

arrhythmias, the condition has been called the syndrome of alternating bradycardia and tachycardia (22), the bradycardia-tachycardia syndrome (23), or the brady-tachy syndrome (24). Used in this review, the term sick sinus syndrome (SSS) is defined as intrinsic inadequacy of the S-A node to perform its pacemaking function due to a disorder of automaticity and/or inability of the S-A node to transmit its impulse to the rest of the atrium. Dysfunction of the S-A node may be accompanied by dysfunction in other parts of the conducting system. Automaticity of lower pacemakers is often suppressed in the SSS and conduction disturbances are common in the atria, A-V node, bundle branches, and ventricles. Interestingly, ventricular ectopic activity is rarely seen in patients with the sick sinus syndrome.

In this review, we emphasize recent information on human sinus node function gained from clinical, electrophysiologic, and pharmacologic studies. Studies of the physiology and pharmacology of the sinoatrial node in animal models are, to a large extent, neglected.

ELECTROCARDIOGRAPHIC FINDINGS

It should be clearly understood that the SSS is not a disease with a single etiology and pathogenesis. It is merely a collection of conditions in which the ECG points to sinus node dysfunction. The type of ECG abnormalities reported in the SSS vary from series to series. Among the factors responsible for variance among the series are differences in definition of SSS, patient groups studied, and methods of investigation.

ECG Manifestations of Automatic Dysfunction (P Cell Dysfunction)

The primary function of the sinus node is to initiate the cardiac impulse. The sinus node performs its pacemaking function readily in the absence of nervous or hormonal influences. However, under physiological conditions, the S-A node is modulated carefully by the autonomic nervous system and by hormones such as thyroid hormone and catecholamines of adrenal origin. When sinus bradycardia occurs, its cause can be extrinsic or intrinsic. Extrinsic bradycardia is usually due to excessive traffic over cholinergic fibers in the vagus nerves. Moreover, when cholinesterase activity is reduced in the S-A node, cholinergic activity can be excessive, even though vagal nerve activity is normal and normal amounts of acetylcholine are released in the vicinity of the S-A node.

Sinus bradycardia in the SSS is most often intrinsic. Sinus bradycardia may be so severe as to cause dramatic symptoms; minimum sinus rates between 15 and 40 can occur. Even when severe, sinus bradycardia does fluctuate with body activity and temperature. Occasionally, pacemaker

function is so poor that sinus arrest occurs. Electrocardiographically, sinus arrest is recognized when the P wave is absent from the body surface ECG and intracavitary atrial electrogram for relatively long periods. Briefer periods of similar nature are often called "sinus pauses." After sinus arrest or sinus pauses, sinus P waves do not resume "on time," i.e. when they would be expected if sinus rhythm continued and sinoatrial block were the cause of the pause. To distinguish sinus pause or arrest from sinoatrial block can be difficult or impossible when sinus arrhythmia is pronounced. Sinus arrest can be very prolonged, even permanent. During prolonged sinus arrest, the atria usually fibrillate or are depolarized by retrograde propagation of impulses arising in the A-V junction. Rarely, sinus arrest is accompanied by atrial standstill. When atrial standstill occurs it suggests infiltrative disease involving the atria, e.g. marked amyloid infiltration of the heart. In sinus arrest with atrial standstill, the atria may not even respond to local electrical stimulation applied via catheter electrodes. Sinus arrest, atrial standstill, and atrial inexcitability usually indicate severe electrolyte abnormality, drug toxicity, or the end stage of infiltrative disease. It is notable that even when sinus bradycardia is severe or sinus pauses occur, A-V junctional or ventricular escape rhythms often do not emerge to assume the pacemaker function of the deficient S-A node. Lack of an escape rhythm in the presence of severe sinus bradycardia implies malfunction of lower pacemakers as well.

Presumably, abnormalities of S-A node automaticity are due to destruction or severe malfunction of the P cells in the node. James et al (25) delineated the morphology of cells in the S-A node and summarized the evidence that clusters of P cells are the site of normal pacemaker activity in the mammalian heart. These workers emphasized the tetrahedral shape and pale cytoplasm of P cells, which are notable for sparsity of myofibrils, mitochondria, and sarcoplasmic reticulum. P cells occur in clusters jointly surrounded by the dense collagen stroma of the S-A node. P cells only attach to other P cells or to transitional (T) cells and never to ordinary atrial cells or cells in the specialized internodal tracts. The attachments of P cells to other cells are few in number, and specialized contact points, e.g. desmosomes, are relatively sparse. The tenuous intercellular connections correlate well with the slow, tenuous conduction in the sinus node. Studies of the pathology of the S-A node in patients with SSS are few. A recent study of eight cases by Brownlee, Evans & Shaw (26) provides some insight into the pathogenesis of SSS. There was a lack of significant atherosclerosis in all eight cases. In six cases the P cells were scanty or absent and the node was densely fibrotic; in a seventh case, the atria and S-A node were infiltrated with amyloid. The loss of P cells seen in this study was much greater than that seen as a normal consequence of aging (27). So far, studies of the

pathology of SSS point away from atherosclerosis as a causal factor, although atherosclerosis and SSS frequently coexist (28). The causes of P cell destruction in SSS are yet to be elucidated.

ECG Manifestations of Sinoatrial Block (T Cell Dysfunction)

For the sinus node to be an effective pacemaker for the heart, it is not enough that a cluster of P cells generate an impulse. Once generated, the primary impulse must successfully emerge from the node and be distributed to the rest of the heart. In SSS, this sequence of events can fail to occur. Several ECG patterns bear evidence that an impulse was formed in the S-A node but failed to activate the atria. First, sinoatrial block with Wenckebach's phenomenon is common in SSS. This pattern of block is easy to recognize when sinus arrhythmia is absent, the P-R intervals are constant, and P-P intervals repeatedly show characteristic Wenckebach sequences, e.g. 5:4, 4:3, or 3:2. Marked sinus arrhythmia can obscure the pattern and infrequent block can simulate sinus arrhythmia. Since sinus and A-V node dysfunction often coexist, the P-R interval may lengthen as the P-P intervals shorten in the Wenckebach structure; occasionally, periodic A-V nodal block may be caused by S-A Wenckebach—so-called double Wenckebach. A second form of sinoatrial block lacks the Wenckebach periodicity and is analogous to Mobitz type II A-V block. Suddenly one or more P waves may be missing, but P waves resume as if the sinus rhythm had never been disturbed. The presumption, then, is that sinus impulses are generated but fail to propagate to the atrium. First-degree S-A block, the analogy of simple P-R prolongation in A-V block, cannot be diagnosed from the 12-lead ECG.

As noted above, connections between sinus node P cells and atrial cells have not been identified by microscopists. Evidently, the only route by which the cardiac impulse can reach the atrium from its site of origin in the S-A node is via the transitional (T) cells. Presumably, the T cells seen in ultrastructural studies (25) correspond to the "perinodal cells" identified by microelectrode studies in rabbit atria (29). These cells form the bridge between P cells and specialized atrial tracts at the margin of the sinus node. The perinodal fibers have electrical properties reminiscent of N cells in the A-V node, e.g. marked time dependence of recovery from the refractory state. In normal circumstances, the sinus impulse is amplified in the perinodal fibers as it conducts toward the atrium. Also, perinodal fibers protect the S-A node from ectopic atrial impulses. The electrophysiologic properties of the perinodal fibers make them likely sites of block and reentry in abnormal circumstances. Experimentally, it has been shown that sinus exit block occurs in the perinodal fibers during digitalis intoxication (30) and

that entrance block and reentry of premature atrial impulses can occur in perinodal fibers (31). Striking histologic changes have been found in post-mortem studies of hearts from patients who showed sinus node dysfunction and/or atrial arrhythmias in life (32, 33). Thus, the ECG findings of S-A block are in accord with information currently available from studies of human pathology and from studies in animal models.

Tachyarrhythmias in the Sick Sinus Syndrome

THE BRADYCARDIA-TACHYCARDIA SYNDROME The bradycardia-tachycardia syndrome is a common form of the SSS. Sinus bradycardia is punctuated by bouts of supraventricular tachycardias. Often several forms of supraventricular tachyarrhythmia are observed over a short period of time in the same patient—atrial fibrillation, atrial flutter, and atrial tachycardia. Also, it is common for bouts of atrial tachyarrhythmias to be of brief duration. When a tachyarrhythmia ceases, there may be a long pause before sinus rhythm or some form of escape rhythm resumes. In flagrant forms of the bradycardia-tachycardia syndrome, the patient may switch from severe sinus bradycardia to an atrial tachyarrhythmia and back many times in the course of a few hours. The posttachycardia pauses may be frighteningly long and the most prominent symptoms may be produced by these pauses. However, symptoms may be prominent during tachycardia as well. It is well to note that some patients with severe bradycardia-tachycardia syndromes have no significant symptoms.

ATRIAL FIBRILLATION Atrial fibrillation can be a default rhythm when sinus node dysfunction is severe, particularly when atrial enlargement and fibrosis accompany sinus node dysfunction. At present, there is no certain way to distinguish default atrial fibrillation in sinus node dysfunction from atrial fibrillation in patients with normal sinus node function. This distinction has practical importance since treatment of atrial fibrillation either with drugs (22) or by cardioversion (15) can have disastrous consequences if the sinus node is inadequate. One clue of severe sinus node dysfunction in patients with atrial fibrillation is a slow ventricular rate in the absence of drugs such as digitalis or propranolol; A-V nodal function is abnormal in more than half of the patients with severe sinus node dysfunction (34, 35). Conversely, when A-V node disease is present, there is an increased likelihood of sinus node disease. Although the probability of sinus disease given A-V node dysfunction is not known precisely, this conditional probability is strong enough to warrant extreme caution when attempting to convert atrial fibrillation with slow ventricular rate to sinus rhythm either by drug treatment or by cardioversion.

Detection of Rhythm Disturbances in the Sick Sinus Syndrome

The diagnosis of SSS is an electrocardiographic one (Table 1). Although the ECG abnormalities of SSS may be easy to detect by standard electrocardiography, the standard ECG is often diagnostically inadequate when the ECG manifestations are intermittent. Prolonged (12–24 hour) ECG recording is a very useful technique in this circumstance. Although the utility of Holter recording for detecting SSS is widely accepted, very few studies actually address this point. Goldberg et al (36) recorded 24 hours of ECG in 130 ambulant patients who complained of syncope, dizziness, or palpitations but who had no arrhythmias detected by the standard ECG. Of the 130 patients, only 16 (12%) did not have an arrhythmia and 96 (74%) had atrial arrhythmias. Pauses in rhythm were found in 16 patients (20%) and supraventricular tachyarrhythmias in 57 (44%). The tape recording was repeated in 84 patients and revealed that 26 had arrhythmias in the second recording that were not present in the first. Lipski et al (37) studied 56 patients who had symptoms suggestive of cardiac arrhythmia and found significant arrhythmias in 30 (55%). Of these 30 patients, 26 (87%) had SSS. Reiffel et al (38) recorded 12 hours of ECG in 51 patients referred because of sinus bradycardia and suspected SSS; only eight were known to have SSS. An extensive workup revealed 28 cases of SSS, 20 (71%) were identified by a single 12-hour recording. Crook et al (39) recorded 72 hours of ECG in 25 consecutive patients with previously diagnosed SSS. Their main finding was that 20 patients (80%) had a previously unnoticed significant arrhythmia found in the prolonged ECG recording.

A much lower incidence of SSS is found when asymptomatic patients are studied or when ECG recording is used to evaluate chest pain, functional capacity, or drug efficacy (40–42). Brodsky et al (43) made a 24-hour ambulatory ECG recording in 50 white male medical students who had no symptoms or cardiovascular disease. They found that 12 (24%) had severe

Table 1 Electrocardiographic patterns of the sick sinus syndrome

-
-
- | | |
|----|--|
| 1. | Long pauses in sinus rhythm caused by sinus arrest, third-degree S-A block, or repetitive concealed sinus exit block. |
| 2. | Periods of second-degree S-A block, inappropriate for drug therapy. |
| 3. | The bradycardia-tachycardia syndromes. |
| 4. | Chronic atrial fibrillation with a slow ventricular rate in the absence of drugs that impair A-V conduction, and inability to resume stable sinus rhythm after electrical cardioversion. |
| 5. | Severe sinus bradycardia, not induced by drug excess. |
-

sinus bradycardia ($< 40/\text{min}$), 14 (28%) had sinus pauses (> 1.75 sec), and 25 (50%) had marked sinus arrhythmia ($> 100\%$ difference in adjacent cycles). These findings may reflect the vagotonia of youth and cannot be used as a control group for elderly patients. Hinkle et al (41) recorded six hours of ECG during standardized activity in a random sample of 283 middle-aged men (median age 58 years). They found that 4 (1.4%) had S-A block and 50 (17.7%) had one or more complex supraventricular dysrhythmias. These authors used the term "sustained relative bradycardia" to describe individuals whose heart rate failed to rise in the afternoon and had a low peak response to exercise. Patients with sustained relative bradycardia had slightly reduced resting heart rates and failed to show the normal abrupt slowing of heart rate and variation due to respiration that characterized the normal subjects. Subjects with sustained relative bradycardia were thought to have disorders of the S-A node and, on follow-up, had an increased death rate. There have been no population studies of cardiac rhythm in the age range where SSS usually occurs. Therefore, we know neither the heart rate norms nor the prevalence of SSS in asymptomatic patients, aged 65 to 80.

If two 24-hour ECG recordings fail to reveal ECG findings of SSS and symptoms are severe but very intermittent, there is a significant likelihood that sinus dysfunction is severe but intermittent. In this circumstance, the patient can carry a pocket-sized device for ECG via telephone. When symptoms occur, the patient or a companion can transmit the ECG to a medical facility using any ordinary telephone. To successfully accomplish transmission, the arrhythmia must persist for about a minute or longer. It is economically feasible for the patient to carry such a device for several weeks. A similar approach is to have the patient wear a recorder which he can activate when he experiences symptoms. This approach can be useful but is more expensive than telephone transmission.

CLINICAL FEATURES OF THE SICK SINUS SYNDROME

The symptoms in patients with SSS may be intermittent, varied, and difficult to associate with ECG changes (23, 24, 44). The symptomatology varies considerably, according to the state of other organ systems, e.g. brain, heart, and kidneys, as well as the type and severity of rhythm disturbance in the SSS. In reviewing published reports of SSS, one is struck by variance in frequency of common symptoms such as syncope, dizziness, or palpitation. The frequency of such symptoms is strongly influenced by the local criteria used to diagnose the SSS. Often symptoms such as dizziness or syncope are used either for case finding or as an actual criterion for the diagnosis of SSS.

For example, a patient whose heart rate is persistently below 50/min and who has a history of dizziness or syncope may be labeled SSS, while a patient with an identical heart rate who has no symptoms is not. Symptoms that influence diagnostic decisions in this manner will almost certainly occur with high frequency in the investigator's series. If the diagnosis of SSS is based solely on ECG abnormalities, patients may have comparable degrees of sinus node dysfunction but a great variability in severity of symptoms. Patients who have severe cerebrovascular, cardiac, or renal disease may have severe symptoms with bradycardia or tachycardia while those without associated disease do not.

In considering reports of SSS, it is well to recognize that, at present, both diagnosis and therapy are often strongly influenced by symptomatology. With this fact in mind, we can compare the frequency of cardinal symptoms in the reports of Moss & Davis (24) and Rubenstein et al (23); patients were carefully characterized clinically in both series. Syncope occurred in approximately half the patients in each series (Rubenstein et al, 25 of 56; Moss & Davis, 11 of 27). Lightheadedness occurred in 27% of the 56 patients in the group reported by Rubenstein et al and in 7% of the 27 patients studied by Moss & Davis; palpitation occurred in 30% (17 of 56) of the former group and 15% of the latter (4 of 27). Increased congestive failure or angina was attributed to SSS in 14% of Moss & Davis's group. It is notable that 10 patients in Rubenstein's series had cerebrovascular accidents; 8 of the 10 were among the 33 patients with bradycardia-tachycardia syndrome. Thromboembolism occurred frequently in patients with SSS, 8 of the 10 cerebrovascular accidents were attributed to emboli and two patients had surgical removal of peripheral emboli. This increased risk of embolism in the bradycardia-tachycardia syndrome is presumed to be due to changing atrial rhythms (37, 45). Ferrer (44) suggests that mildly reduced cerebral blood flow can produce such protean symptoms as irritability, nocturnal wakefulness, memory loss, and errors of judgement. The patient may be less aware of such changes than his family or friends and therefore may give a negative history. The patient's family and his physician may attribute such vague changes to senility.

If coronary perfusion or ventricular function is reduced, the rate and rhythm problems of the SSS may cause further deterioration. These may be dramatic events such as acute pulmonary edema or angina pectoris. However, the cardiac functional deficit of SSS may be as subtle as fatigue, mild digestive disturbance, periodic oliguria or edema, and mild intermittent dyspnea (44). It may be very difficult to ascertain the extent to which the SSS contributes to cardiac signs and symptoms of coexistent cardiac disease. Ferrer emphasized the clue offered by finding an inappropriately slow heart rate in circumstances such as febrile episodes or acute left ventricular failure.

ELECTROPHYSIOLOGIC TESTS OF SINUS NODE FUNCTION

Over the past seven years, a great deal of effort has been expended in developing and evaluating clinical electrophysiologic testing of sinus node function. These tests are particularly useful if repeated and/or prolonged ECGs fail to incriminate the sinus node when the level of clinical suspicion is high. Examples of conditions that may elude observation during ECG recording but produce clearcut abnormalities during clinical electrophysiologic testing are intermittent S-A arrest, intermittent S-A block, and S-A block in patients with marked sinus arrhythmia. Electrophysiologic tests can determine whether patients who recover from second-degree S-A block during intercurrent illness or drug therapy have completely recovered or have residual first-degree S-A block. In addition to their use in evaluating suspected SSS, functional tests are useful for evaluating the effects of cardioactive drugs on normal and abnormal sinus nodes. The two most widely used clinical electrophysiologic tests are atrial overdrive pacing and premature atrial stimulation.

Atrial Overdrive Pacing

CORRECTED SINUS RECOVERY TIME The response of pacemaker cells to overdrive stimulation is transient suppression of automaticity (46, 47). In the normal S-A node, overdrive suppression is thought to result from a transient increase in membrane potassium conductance of P cells due either to a local increase in extracellular K^+ or to enhanced release of acetylcholine. Under the hypothesis that recovery after pacing would be excessively delayed when S-A node automaticity was defective, clinical electrophysiologists recently began to measure sinoatrial recovery time (SRT) (48–56). When atrial pacing is stopped, a pause occurs, then sinus rhythm resumes (see Figure 1); the postpacing pause is called the SRT. To reduce variability in SRT caused by between-patient differences in heart rate, most investiga-

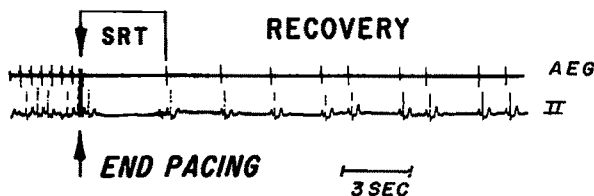


Figure 1 Determination of the sinus recovery time (SRT) by overdrive atrial pacing. The top trace is a bipolar atrial electrogram (AEG) and the lower is lead II of the electrocardiogram. At the arrows, atrial pacing is discontinued. The SRT is the time from the last paced atrial depolarization to the first spontaneous atrial depolarization. The CSRT is obtained by subtracting the spontaneous atrial cycle length from the SRT. Note the secondary sinus pauses.

tors "correct" the SRT, usually by subtracting the prepacing sinus cycle length (SCL), i.e. corrected sinus recovery time, $CSRT = SRT - SCL$ (50, 52, 53, 56). A prolonged CSRT is used as an index of abnormal S-A node automaticity.

Several additional factors should be considered when interpreting the CSRT (57). First, 30 or more seconds of overdrive are required in order to establish a steady-state of sinus suppression (38, 48, 53, 55). Second, the proximity and accessibility of the pacing site to the S-A node can influence the CSRT (57). Most, but not all, laboratories measure CSRT with the pacing electrode against the mid-lateral right atrial wall and the recording electrode near the junction of the superior vena cava and right atrium. Third, the CSRT obtained during clinical testing is measured from atrial rather than S-A node depolarization. Therefore, CSRT measures not only the increment due to overdrive suppression but also includes two conduction times: (a) the time from the last paced atrial depolarization to the S-A node, and (b) the time from the postpacing S-A node escape depolarization to its atrial manifestation. Thus, CSRT may be prolonged not only by excessive suppression of automaticity but also by intraatrial or perinodal fiber conduction delay. Hence, CSRT does not simply reflect S-A node automaticity. Fourth, the magnitude and reproducibility of SRT or CSRT measurement depends on the pacing rate (38, 48, 49, 55, 56). In normal subjects, Mandel and co-workers (48) showed that as the pacing rate is increased the CSRT will increase until the peak CSRT is reached at about 130/min. Pacing faster than 130/min often produces shorter CSRTs. In patients with sick sinus syndrome or sinus bradycardia, the pacing rate that yields the maximal CSRT is unpredictable but is frequently lower than in normal subjects, often in the range of 70–90/min (58) (see Figure 2). Therefore, pacing must be performed at multiple rates from 60 to about 150/min in order to determine the maximum CSRT (see Figure 2). Also, CSRT can vary when it is measured repeatedly at the same pacing rate even though the catheter position, duration of pacing, and time between measurements are held constant. Such variation tends to be greatest in patients who have prolonged CSRT values. Variation among CSRT values at one pacing rate is most striking when pacing is faster than the rate that produces the maximum CSRT. The variation in CSRT seen at higher pacing rates is probably caused by rate dependent conduction delays in the perinodal fibers and S-A node. Since the sinus perinodal zone behaves in a manner analogous to the A-V node, it is reasonable to attribute differences among replicate measures of CSRT to variable perinodal conduction delay, even though the recovery time in the S-A node itself is constant in each stimulation sequence. In some patients with SSS, overdrive suppression cannot be achieved by atrial pacing because A-S conduction block occurs at low atrial pacing rates. In such patients, the CSRT may prolong after administration

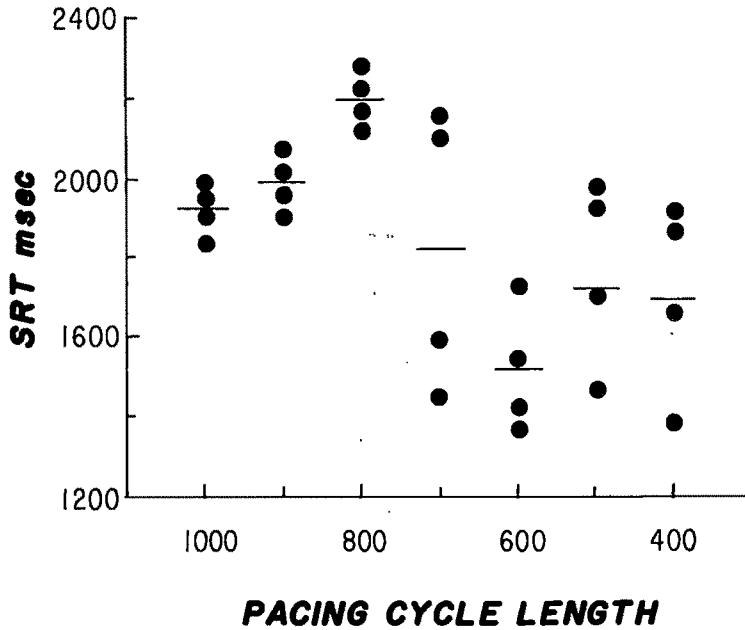


Figure 2 The graph shows the SRT determined four times at pacing cycle lengths between 1000 and 400 msec. The patient's spontaneous sinus cycle length was 1330 msec. At a pacing cycle length of 800 msec, the maximum SRT of 2230 msec was obtained (CSRT = 900 msec). At shorter pacing cycle lengths, the mean SRT fell and greater variance developed among replicate measures.

of atropine, a drug that improves conduction in the S-A junction (59-61). Also by analogy to the A-V node, the minimum pacing rate at which atriosinus conduction disturbances occur is lower in patients with abnormalities in S-A conduction than in normal subjects. The pacing rate that yields the longest mean CSRT is not only dependent on S-A conduction but also on spontaneous sinus cycle length. Fifth, the SRT may be shorter than the SCL, causing negative values for CSRT; this event may be caused by sinoatrial echoes and/or S-A entrance block. Similarly, the CSRT may be prolonged due to repetitive concealed reentry in the S-A junction rather than abnormally suppressed automaticity (31). Whether a prolonged CSRT reflects predominantly a defect in automaticity or conduction or both is not very important clinically if the objective is to distinguish normal from abnormal sinus node function. To date, all studies agree that the majority of patients with sinus node dysfunction have prolonged CSRTs but unfortunately a significant fraction do not, probably as a result of S-A entrance block during pacing, intermittency of dysfunction, or S-A echoes.

SECONDARY SINUS PAUSES AND THE "ENTIRE SINUS RECOVERY TIME" Abrupt lengthening of postpacing atrial cycles after the first escape cycle was noted in 1973 and called secondary sinus pauses (51) (see Figure 1). Secondary sinus pauses are usually observed in patients with sinus node dysfunction and S-A conduction delay and are therefore attributed to conduction abnormalities in the S-A junctional tissues (51, 55, 62). Atropine treatment usually abolishes secondary sinus pauses by improving conduction in perinodal tissues. Recently, several authors suggested that secondary sinus pauses are useful for identifying sinus node dysfunction and may be a more sensitive indicator than the CSRT (55, 62). Delius & Wirtzfeld (55) report that, even in normal subjects, overdrive suppression persists beyond the first postpacing atrial cycle. These workers suggest that the time to reestablish the prepacing sinus rate ("entire SRT") may be of value for evaluating sinus node function. Like the CSRT, the entire SRT may reflect abnormalities of either S-A node automaticity or S-A conduction. Delius & Wirtzfeld found that the entire SRT averaged 3.8 sec in their control group and 8.3 sec in patients with SSS (55). However, as with the CSRT, there was overlap between normals and SSS. Further investigation of secondary pauses and the entire SRT will be necessary in order to establish their usefulness in detecting S-A node dysfunction.

Premature Atrial Stimulation

Programmed premature atrial stimulation has also been widely used to test S-A node function, principally to estimate sinoatrial conduction time (SACT) (38, 54, 56, 63-67). The technique consists of evoking atrial premature depolarizations (APDs) with stimuli introduced into selected portions of the spontaneous atrial cycle (see Figure 3). Responses to the APDs are described by the following terminology: A1A1 is the interval between the two sinus P waves immediately preceding the APD; A1A2 is the interval between the last sinus P wave and the APD; A2A3 is the interval between the APD and the subsequent sinus P wave (the return cycle); and A3A4 is the sinus cycle immediately following the return cycle. To evaluate S-A node function with premature atrial stimulation, all of atrial diastole is scanned several times at 10-20 msec intervals using a programmed stimulator. To summarize and display the results, the A2A3 and A3A4 intervals are plotted as a function of A1A2 intervals as in Figure 3. For purposes of discussion the summary plot of premature atrial stimulation is divided into three zones.

ZONE I In the last part of the atrial cycle (approximately 0.75-1.00 of the cycle), A2s propagate toward the S-A node but collide with the emerging sinus impulse. Since the S-A node is not reset, A2A3 is fully compensatory,

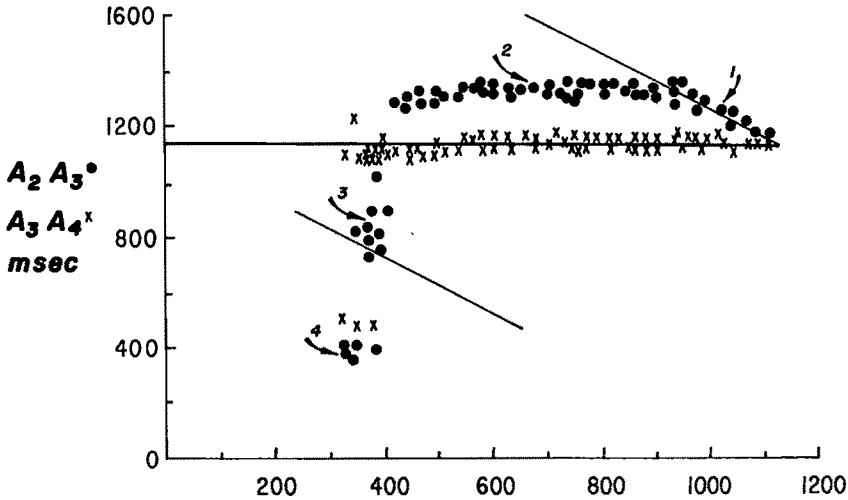


Figure 3 Premature atrial stimulation. The A2A3 (•) and A3A4 (x) plotted as a function of A1A2. Here 1 = interference, 2 = reset, 3 = interpolation, 4 = S-A echoes. The SACT is estimated from measurements obtained in the zone of reset.

i.e. $A1A2 + A2A3 = 2(A1A1)$. In some patients A2A3 intervals are unexpectedly long so that $A1A2 + A2A3$ is greater than $2(A1A1)$. This response is usually seen in patients with marked sinus arrhythmia.

ZONE II. In the middle of atrial diastole (about 0.30–0.75 of the cycle), A2s penetrate and depolarize the S-A node before it spontaneously reaches threshold. When sinus reset occurs, three components contribute to A2A3: (a) conduction time from atrium to S-A node pacemaker site, (b) the time between the sinus response to A2 and the next spontaneous sinus action potential, and (c) conduction time from S-A pacemaker site to the atrium. In normal human subjects with minimal sinus arrhythmia, A2A3 remains quite constant over a very long portion of atrial diastole, which indicates that the sum of these three intervals is relatively constant. When sinus arrhythmia is pronounced, A2A3 will vary considerably, but even so $A2A3 - A1A1$ will be relatively constant, which indicates that the cause for variability in A2A3 is the changing sinus cycle length. Thus, the sum of conduction time into and out of the S-A node is reasonably constant even during sinus arrhythmia. When the responses in Zone II are stable over a long segment of atrial diastole, SACT can be estimated by $\frac{1}{2}(A2A3 - A1A1)$. This calculation of SACT assumes that (a) sinus node automaticity is not importantly disturbed by A2, (b) sinus reset occurs, and (c) conduction times into and out of the S-A node are similar. It is important to

consider these assumptions when examining the summary plot. If sinus automaticity is significantly perturbed by A2, then A3A4 is likely to be longer than A1A1. Careful examination of A3A4 responses is very helpful for detecting depression of automaticity (64, 68). Reset may fail to occur at either extremity of Zone II. At the junction of Zones I and II, electrotonic interaction of A2 with the S-A node may abbreviate A2A3 and mimic sinus reset (69). Early in Zone II, A2 may block on the rim of the S-A node and A2A3 will be shorter than during reset (see Zone III below). SACT is most meaningfully estimated from mid Zone II in a segment where the A2A3 and A3A4 are constant; the appropriate responses can easily be identified from the summary plot.

ZONE III In some human subjects, A2A3 remains constant as early as A2 can be elicited so that only Zone I and II responses are obtained. However, a third zone may be seen early in atrial diastole (about 0.20–0.40 of the cycle) in which a number of complex phenomena may occur (51). In Zone III, A2 may block in the perinodal tissues and produce one of three main effects on the emerging sinus impulse (70, 71). If the emerging sinus impulse conducts normally, $A1A2 + A2A3$ will approximately equal A1A1, i.e. A2 will be precisely interpolated (see Figure 3). If the perinodal tissues are relatively refractory when the sinus impulse emerges, conduction delay will result and $A1A2 + A2A3$ will be greater than A1A1 but less than A2A3 in Zone II. If perinodal tissues are absolutely refractory, the emerging sinus impulse will block in the perinodal tissue, producing an $A1A2 + A2A3$ that is $2(A1A1)$. In addition to the effects on the subsequent sinus impulse, an early A2 can conduct so slowly in the perimeter of the S-A node that it reexcites the atrium to produce a sinus node echo (72–76) (see Figure 3). A2 may produce recirculating activity in the vicinity of the S-A node that tends to bursts of supraventricular tachycardia (77, 78). Supraventricular tachycardia caused by sinus node reentry is usually unstable and of short duration.

FIRST-DEGREE S-A BLOCK Premature atrial stimulation can reveal the presence of first-degree S-A block in man; in the absence of spontaneous atrial parasystole, this diagnosis cannot be made in any other way (51, 63). First-degree S-A block is diagnosed when the SACT is very long or when Zone I extends through all of atrial diastole. The latter response indicates sinus entrance block throughout the cardiac cycle and is presumed to indicate sinus exit block as well. This assumption is supported by the strong association between the diagnosis of first-degree S-A block by premature atrial stimulation and intermittently occurring second-degree S-A block. We use this test to evaluate patients who had second-degree S-A block

during either an acute illness or treatment with drugs with the potential to depress sinus node function. The presence of first-degree S-A block during stimulation indicates significant residual or covert dysfunction. A normal SACT, under these circumstances, suggests either good recovery of the node or intermittent malfunction.

DIAGNOSTIC USE OF THE SINOATRIAL CONDUCTION TIME There have been objections to the calculated SACT on the grounds that premature atrial stimulation may suppress S-A node automaticity or that the two conduction times comprising the interval may be unequal. Even when these objections hold, SACT is useful in identifying individuals with sinus node dysfunction. Although many patients with sinus dysfunction have normal values for SACT (38, 54, 56, 62), the use of SACT along with CSRT augments the diagnostic sensitivity of clinical electrophysiologic testing of S-A node function (38, 56, 67). In subjects with normal heart rate and sinus function, the SACT is usually between 60 and 80 msec (35, 38, 51, 65, 67). There has been disagreement about the upper limit of normal for the SACT; it seems to lie between 120 and 150 msec in subjects with normal heart rates or with mild bradycardia (38, 54, 56, 67, 79). The relationship between heart rate and SACT has not been fully elucidated. In individual patients with sinus arrhythmia there is an inverse relationship between instantaneous atrial cycle length and SACT (80). On the other hand, for groups of patients, no relationship has been established between sinus rate and SACT. In the final analysis, the SACT is useful for detecting sinus node dysfunction because it can be clearly abnormal when the CSRT is normal; in fact, there is little correlation between the CSRT and SACT in SSS (35, 54).

Digitalis

The safety of digitalis therapy in patients who have a significant sinus node dysfunction is an unsettled question. Some authors state that digitalis is safe in the sick sinus syndrome while others warn of its hazards (21, 23, 81-84). To obtain further information on this issue, several investigators have performed electrophysiologic studies before and after acute administration of digitalis (84-91) (see Table 2). Some studied subjects who were assumed to have had normal sinus node function (86, 88) while others studied patients who had persistent sinus bradycardia or clearcut sick sinus syndrome (84, 85, 89-91). All studies found that digitalis usually had little effect on the sinus rate either in normal subjects or those with SSS. In no study was there a significant change in mean heart rate. For individual patients, heart rate is almost as likely to increase as to decrease and the magnitude of the change is small, rarely more than five beats per min (see Table 2).

Table 2 Effect of digitalis on sinus node function

Author (ref.)	Digitalis preparation (mg/kg)	Subjects (no.)	SCL (msec)		CSRT (msec)		SACT (msec)	
			Before	After	Before	After	Before	After
Engel et al (84)	Ouabain 0.01	14	1,043	1,060	1,133	693	—	—
Cummings et al (91)	Ouabain 0.015	14	1,079	1,010	526	251	—	—
Vera et al (90)	Digoxin 0.04	9	714	565	714	565	—	—
Bond et al (86)	Ouabain 0.01	5 ^a	—	—	—	—	—	27 ^b
Dhingra et al (88)	Ouabain 0.01	16 ^a	887	905	226	103	90	110
Reiffel et al (89)	Digoxin 0.75	18	1,011	1,044	623	350	105	108

^aNormal subjects.^bIncreased by 27 msec; before and after values not given.

A striking and unexpected finding was the decrease in CSRT seen after digitalis administration. A decrease in CSRT was seen both in normal subjects and those with SSS after administration of either ouabain or digoxin (84, 89–91). Often, the magnitude of decrease in CSRT is quite large and potentially of clinical significance. Cummings et al (91) reported a patient who had postpacing pauses of 10.7 sec before ouabain and decreased to 1.4 sec after 0.015 mg/kg of ouabain. The electrophysiologic findings seem to correlate with spontaneous events in patients with the SSS. Cummings et al (91) noted no deleterious effects of ouabain in 12–72 hours of postouabain ECG recording in 14 patients with SSS. One patient had had spontaneous sinus pauses of up to 7.0 sec prior to ouabain administration and was quite symptomatic during these pauses; after ouabain, his spontaneous sinus pauses disappeared. However, one cannot give digitalis to SSS patients with impunity. Margolis et al (87) reported prolonged (4.0 sec) postpacing pauses apparently caused by digoxin in a 53-year-old American Indian who was studied after mitral valve replacement for mitral valve prolapse. Without digoxin, their patient had a normal CSRT and SACT but failed to increase his heart rate after intravenous atropine. Although this case is a very unusual one, it indicates the need for caution when giving digitalis to patients with sinus node dysfunction.

The response of SACT to digitalis administration has not been studied as extensively as the heart rate and CSRT responses. Dhingra et al (88) studied 16 and Bond et al (86) five subjects with apparently normal sinus node function before and after ouabain administration. Both groups found an increase in SACT after ouabain of about the same magnitude, 20 and 27 msec respectively. We gave 0.75 mg of digoxin to 18 subjects who had sinus bradycardia; seven of these had clearcut SSS (89). We found no change in the group mean for SACT; the largest increase in SACT was 37 msec. Six of the seven patients with SSS showed some increase in SACT and one patient developed second-degree S-A block. Bond et al (86) also studied

two patients with recent recoveries from second-degree S-A block attributed to digitalis administration. Second degree S-A block developed in both patients when they were challenged with ouabain. It seems likely that digitalis can cause clinically significant depression of S-A conduction in a minority of patients with SSS.

The mechanism of digitalis action on the human sinus node is incompletely understood. After vagal blockade with atropine, digoxin always slows the heart rate, prolongs the CSRT, and usually prolongs the SACT (89). Since these effects are in the same direction as those produced by vagal influence, it seems paradoxical that digoxin often does not decrease the heart rate or increase the CSRT or SACT before vagal blockade. Digitalis uniformly slows the sinus rate in other mammals with intact vagi (92).

Several observers recommend that all patients with SSS who require digitalis therapy first be treated with a permanent electronic pacemaker (21, 81, 82). However, data presented above and the clinical experience reported by Rubenstein et al (23) suggest that most patients with SSS can safely take digitalis. Shortening of the CSRT by digitalis may be beneficial in the bradycardia-tachycardia syndrome; the pause following spontaneous tachyarrhythmias may be shortened. The tendency of digitalis to prolong SACT could produce or aggravate S-A block but this usually does not occur. Our experience and that of Engel et al (84, 85) suggest that the results of electrophysiological testing following acute digitalis administration will predict the long-term effects of the drug. Patients who have favorable responses during acute electrophysiologic testing have done well during follow-up. However, the relationship between the acute effects of digitalis on sinus node function and the response to long-term treatment is an important issue that should be studied further. If electrophysiologic studies are equivocal, the patient may be treated with digitalis for a longer period of time with a temporary demand pacemaker in place. The final course of action can then be judged on the basis of prolonged ECG observation. Patients who show marked detrimental effects of digoxin during electrophysiologic testing should be treated with implanted pacemakers if digitalis must be given chronically.

Antiarrhythmic Drugs

QUINIDINE AND PROCAINAMIDE In 1954, D. H. Short (22) warned of the dramatic adverse effects of quinidine on sinus node function in his description of four patients with bradycardia-tachycardia syndrome. The effect of quinidine and procainamide on sinus rate is difficult to predict because their direct suppressive action on the S-A node may be counteracted by their autonomic effects. The autonomic effects of quinidine are complex: It exerts mild to moderate cholinergic and alpha adrenergic block-

ade. Cholinergic blockade can raise the heart rate and block vagal reflexes. Alpha blockade can blunt the vasoconstriction that usually occurs during the baroreceptor reflex and can thus intensify the traffic on the cardiac sympathetic nerves. Procainamide does not produce significant alpha blockade but has been shown to produce mild ganglionic blockade in the dog (93).

Neither quinidine nor procainamide usually has much effect on sinus rate in normal human subjects. In fact, heart rate remains remarkably stable even during intravenous administration of procainamide. However, occasional patients develop marked sinus bradycardia or other changes of sick sinus syndrome following quinidine or procainamide administration (95–99). Josephson et al (96) studied the effects of intramuscular quinidine on 21 subjects and found, on the average, a small increase in heart rate. Since the systolic blood pressure fell, the increased heart rate may have been due to autonomic rather than direct effects. Mason et al (94) studied the effect of quinidine on four transplanted human hearts and found, following quinidine, that the donor S-A node always slowed while the recipient S-A node accelerated. This suggests that quinidine has a significant vagolytic and/or sympathomimetic effect on the innervated human sinus node as well as a direct suppressive effect. Patients with minimal vagotonia are more likely to develop a decrease in heart rate during quinidine treatment, because the direct suppressive effect is unopposed. Similarly, quinidine or procainamide treatment is more likely to aggravate sinus node dysfunction in patients with atropine-resistant SSS. Calon et al (100) measured the CSRT in 17 patients who were beginning oral quinidine therapy. They found no effect in five patients with normal S-A node function, but CSRT prolonged very significantly in those with paroxysmal atrial fibrillation (nine patients) or SSS (three patients). Each of the three patients with SSS became symptomatic as a result of very long postpacing pauses (100). Recently, Wyse et al (101) performed electrophysiologic studies in 16 patients before and during procainamide treatment. During procainamide infusion, the sinus cycle length decreased slightly and the CSRT and SACT shortened significantly.

DISOPYRAMIDE PHOSPHATE Disopyramide phosphate is a new antiarrhythmic drug electrophysiologically similar to quinidine and procainamide but with a more pronounced anticholinergic effect (102–105). Like quinidine and procainamide, disopyramide has little effect on sinus rate in normal subjects (102, 103). Unlike quinidine, disopyramide shortens the CSRT (102). To our knowledge, there has not been any case report of S-A node dysfunction induced by this agent. Strauss et al (104) reported preliminary results of electrophysiologic testing in seven subjects; sinus cycle length and

CSRT either decreased or remained unchanged. If these observations are confirmed, disopyramide may be the safest quinidine-like antiarrhythmic agent to use in patients with sick sinus syndrome and ectopic activity. For the moment, disopyramide should be used with caution in patients with SSS, since toxic concentrations may depress sinoatrial automaticity (105).

PROPRANOLOL After therapeutic doses of propranolol, resting sinus rate decreases about 10–20% in normal subjects (106–110). However, severe sinus bradycardia, sinoatrial arrest, or sinoatrial exit block occasionally occur (106–109). In their review of beta blockers, Gibson & Sowton (109) speculate that, in normal subjects, prolongation of sinus cycle length is predominately the result of propranolol's beta-blocking effect, while, in patients with sick sinus syndrome, aggravation of sinus node dysfunction is a result of both the beta blocking and direct depression. Strauss et al (110) studied 10 patients with symptomatic S-A node dysfunction and found that propranolol increased the average sinus cycle length, CSRT, and SACT. A few of their patients showed marked S-A node suppression.

LIDOCAINE AND DIPHENYLHYDANTOIN Lidocaine given alone (111–115) or in combination with other drugs (115–117) can depress S-A node function. The prevalence of severe sinus node depression by lidocaine is unknown but is probably quite low. In vitro (119) and in vivo (120) animal experiments suggest that lidocaine suppresses sinoatrial automaticity and conduction only at toxic concentrations in the normal mammalian heart. However, significant S-A node dysfunction can occur after even small intravenous doses in diseased human hearts (111). Dhingra et al (121) studied the effect of intravenous lidocaine on S-A node function in 13 patients with normal S-A node function and in 14 patients with S-A node dysfunction. Lidocaine had no effect on sinus cycle length or on CSRT in either group. SACT was unaffected in normal subjects, but prolonged by about 10 msec during lidocaine infusion in patients with S-A node dysfunction (121).

Like lidocaine, diphenylhydantoin (DPH) has been reported to cause sinus node dysfunction in man (122), although there are fewer reports than for lidocaine. DPH has a variable effect on sinus rate, which most often remains unchanged (30, 123–125). Animal studies show that therapeutic concentrations of DPH have no significant direct effects on phase 4 of the normal S-A node P cells (118), which suggests that, in normal subjects, changes in sinus rate due to DPH result from reflex alterations or from effects on the central or autonomic nervous systems. DPH may depress sinus node function when preexisting S-A node damage is present (30) or toxic concentrations are reached (122). DPH toxicity has resulted in both

marked sinus bradycardia and sinoatrial arrest (122). Diphenylhydantoin should, therefore, be used cautiously in patients with SSS.

Since antiarrhythmic drugs not only decrease sinus node function but also depress atrial, junctional, and ventricular automaticity, escape rhythms may fail when the S-A node is depressed. The results may be disastrous. Use of antiarrhythmic drugs to treat the tachycardia of bradycardia-tachycardia syndrome, or to treat unrelated arrhythmias, requires careful observation and, not infrequently, concomitant pacemaker therapy.

Lithium

Lithium carbonate, a psychotherapeutic agent used frequently to treat bipolar depression (126, 127), can produce marked sinus bradycardia, sinoatrial arrest, and/or sinoatrial block (128–134). The prevalence of such striking changes is unknown but is probably low. Seven cases of lithium-induced S-A block or S-A pauses are reported in the literature (128–134). Questions arise: Do these cases represent aggravation of preexisting S-A node dysfunction? unusual sensitivity to lithium? excess plasma-lithium concentration? In most reports, plasma lithium was measured and found to be within the usual therapeutic range (0.5–2.0 meq/liter). There are hints in the literature that lesser degrees of sinus node dysfunction may be seen more frequently during lithium therapy. Middelhoff & Paschen (135) found that sinus bradycardia was “frequent” in patients taking lithium. Tilkian et al (136) studied 10 patients with graded treadmill exercise and noted that, at equivalent levels of exercise, sinus rate was slower during lithium treatment, even though there was no evidence of sinus node dysfunction on 24-hour ECG recordings.

Detailed electrophysiological evaluation of the sinus node in patients both on and off lithium has seldom been performed. Wellens et al (131) studied a 56-year-old woman who presented after six months of repeated syncope. On admission she had sinus bradycardia (45/min) and sinus pauses. Her lithium level was normal (1.0 meq/liter). She underwent electrophysiologic studies off lithium and again after lithium was reinstituted. These showed that the CSRT was completely normal off lithium but became markedly prolonged during treatment. Moreover, the study revealed that lithium had no effect on other portions of the conduction system. Wilson et al (132) studied a 64-year-old woman admitted with unsteadiness of gait, marked sinus bradycardia (35/min), and sinus pauses. Her studies showed that she was entirely normal off lithium but the drug caused a marked increase in CSRT and minimal increases in the A-H and H-V intervals of the His bundle electrogram. We studied a third patient who developed sinus node dysfunction on lithium. Prior to treatment, the sinus rate was 55–60 min. During lithium therapy, symptomatic sinoatrial arrest and exit block appeared. Electrophysiological testing (see Table 3) revealed occult sinus

Table 3 Effect of lithium on sinus node function

	SCL	Maximum CSRT	SACT	Atrial FRP	Atrial ERP	A-H	H-V	PCL at Wenckebach
Control	865	727	SAB ₁	478	359	100	50	400
Lithium	940	2,510	SAB ₂	451	340	90	50	400

^aAll measurements in msec. Abbreviations are as follows: FRP = functional refractory period, ERP = effective refractory period, A-H = A-V nodal conduction time, H-V = His Purkinje activation time, PCL = paced cycle length causing A-V Wenckebach, SAB₁ = first-degree sinoatrial block, SAB₂ = second-degree sinoatrial block.

node dysfunction while off drug, which worsened significantly during lithium therapy. The cases of Wellens and Wilson suggest that normal plasma-lithium concentrations can evoke severe but reversible dysfunction in otherwise normal S-A nodes. Our case suggests that preexisting or latent S-A node disease may predispose to dysfunction on lithium.

The mechanism by which lithium depresses sinus node function is uncertain. Atropine improved sinus node function in the case of Wellens et al (131), but this does not mean that lithium induced sinus node dysfunction by augmenting vagal effects. In fact, Basuray & Dutta (137) showed that chloride lithium reduces the effect of vagal stimulation on the sinus rate in the rabbit. Chronic lithium treatment in man reduces the pressor and heart rate effects of infused norepinephrine (138), although acute experiments in rabbits show that lithium has no effect on the S-A node response to stimulation of cardiac sympathetic nerves (139). It is likely that lithium alters sinus node function by changing the electrolyte status of sinus node cells (139–141). It is known that the lithium permeability of resting heart muscle is relatively high and that lithium can pass through the sodium channel during activation (140, 141). However, once inside the myocardial cell, lithium is not extruded and tends to accumulate (140, 141). Carmeliet (141) showed that, when lithium totally replaces sodium in the perfusate of isolated cat ventricular muscle (i.e. $[Li]_o = 150$ mM/liter), intracellular lithium concentration rapidly rises to about 120 mM/liter after an hour of perfusion, while intracellular potassium and sodium fall to 50% and 20% of their normal values, respectively. When mammalian heart muscle is perfused with Tyrode's solution containing 15 mM/liter lithium, the intracellular lithium concentration at steady-state is about 30 mM/liter (141). Nothing is known of the intracellular ionic composition of heart cells after chronic exposure to solutions containing lithium in concentrations near those attained during chronic therapy of mania (0.5–2.0 meq/liter). Studies in animal models show that high extracellular lithium concentrations lead to a decrease of the spontaneous firing rate in both the sinus node (139, 142) and Purkinje fibers (141, 142). Under these conditions, the potassium permeability is increased and the driving force for the background inward current carried by lithium

is reduced (141); both these conditions act to decrease the spontaneous firing rate. Again, there have been no studies that closely simulate the clinical therapeutic situation.

Therapy and Prognosis

Over the past 10 years, implanted pacemakers have played an increasingly important role in the treatment of SSS. Also, as SSS has become more widely recognized and better understood, increasingly more pacemakers are implanted for this condition. In 1975, Montefiore Hospital performed 200 initial pacemaker implants; 37% were for SSS compared to 18% for permanent complete heart block (143). Another 4% of their pacemakers were implanted for atrial fibrillation with a slow ventricular response (less than 50/min). Of the pacemakers implanted for SSS, 55% were for the bradycardia-tachycardia syndrome and 45% for either sinus arrest (31%) or severe sinus bradycardia (14%). In 1976, 45% of the 176 new pacemaker implantations at our institution were for SSS. Parenthetically, it is important to note that transvenous pacemaker insertion is well tolerated by patients with SSS, even though they tend to be very old and may have severe cardiac or renal disease.

Prior to pacemaker therapy it was very difficult to control symptoms in patients with SSS. Drugs were often ineffective and frequently caused inconvenient or intolerable effects. In general, pacemaker therapy provides remarkable relief of symptoms in SSS (144-148). Nearly all workers report excellent results in those patients whose symptoms are attributable to S-A arrest or severe sinus bradycardia. There is some disagreement about the efficacy of pacemaker therapy for the control of symptoms in the bradycardia-tachycardia syndrome. Conde et al (144) implanted permanent ventricular pacemakers in 31 patients with symptomatic bradycardia-tachycardia syndrome who were not relieved by drug therapy. Of these, 27 (87%) had good relief of symptoms with minimal supplementary drug therapy. In four patients, symptoms continued, but the pacemaker proved useful by facilitating drug therapy. Aroesty et al (146) reported relief from arrhythmias and symptoms after pacemaker implantation in 26 of 28 (93%) patients with bradycardia-tachycardia syndrome. Eleven patients with bradycardia-tachycardia syndrome were among the 52 treated by Chokshi et al (145); in seven of these (64%), supraventricular arrhythmias and symptoms were controlled by pacemaker plus drug therapy. Krishnaswami & Geraci (147) found a lower proportion of patients with bradycardia-tachycardia syndrome relieved of symptoms by pacemaker treatment than those who were paced for sinus bradycardia or arrest. There is a high incidence of stroke in patients with bradycardia-tachycardia syndrome treated with a pacemaker (147), presumably due to cerebral emboli. Thus,

pacemaker therapy does not seem to reduce the high likelihood that patients with bradycardia-tachycardia syndrome will experience stroke (45, 149) and these patients should be advised accordingly. Wohl et al (148) found persistent symptoms in 14 of 36 (39%) patients treated by pacemaker for SSS; persistent symptoms were common in patients with sinus bradycardia and arrest as well as those with bradycardia-tachycardia syndrome. Also, Wohl et al found that patients whose symptoms persist have a significantly higher death rate than those whose symptoms abate after pacemaker implantation. Thus, it seems clear that most patients with SSS who have symptoms clearly attributable to bradycardia will have symptomatic relief after pacemaker implantation. In patients with bradycardia-tachycardia syndrome, results are not as good but still quite acceptable.

The mortality rate after implantation is not as strongly related to the severity of the SSS as to the presence and severity of associated cardiac and other diseases. The overall mortality in the year after pacemaker implantation ranges from 15 to 33% (145-148). Fruenhan et al (150) and Wohl et al (148) report a two-year mortality of 50% and 42%, respectively. Several authors have evaluated the cause of death in SSS patients who were treated with implanted pacemakers. Very few patients die of either pacemaker failure or of uncontrolled supraventricular arrhythmias. The major causes of death in Chokshi's series were myocardial infarction, heart failure, stroke, ventricular arrhythmias, renal failure, and sepsis (145). Aroesty et al found that patients who died in the first year after implantation had much more ischemic and valvular heart disease, left ventricular enlargement, and renal disease than those who survived (146). Both Fruenhan et al (150) and Wohl et al (148) underscore the high prevalence of cardiovascular disease in those who die in the first few years after pacemaker implantation.

Literature Cited

1. Keith, A., Flack, M. 1907. Form and nature of the muscular connections between the primary divisions of the vertebrate heart. *J. Anat. Physiol.* 41:172-89
2. Flack, M. 1910. An investigation of the sino-auricular node of the mammalian heart. *J. Physiol.* 51:64-77
3. Eyster, J. A. E., Meek, W. J. 1914. Experiments on the origin and propagation of the impulse in the heart—the point of primary negativity in the mammalian heart and the spread of negativity to other regions. *Heart* 5:119-36
4. Eyster, J. A. E., Meek, W. J. 1916. Experiments on the origin and conduction of the cardiac impulse. VI. Conduction of the excitation from the sino-auricular node to the right auricle and auriculo-ventricular node. *Arch. Intern. Med.* 18:775-99
5. Eyster, J. A. E., Meek, W. J. 1921. The origin and conduction of the heart beat. *Physiol. Rev.* 1:1-43
6. Lewis, T. 1910. Galvanometric curves yielded by cardiac beats generated in various areas of the auricular musculature: Pacemaker of the heart. *Heart* 2:23-46
7. Lewis, T., Oppenheimer, B. S., Oppenheimer, A. 1910-1911. Site of origin of the mammalian heart beat: The pacemaker in the dog. *Heart* 2:147-69
8. Wybauw, R. 1910. Sur le point d'origine de la systole cardiaque dans l'oreillette droite. *Arch. Int. Physiol. Biochim.* 10:78-89

9. Trautwein, W., Zink, K. 1952. Über membran- und aktionspotentiale einzelner myocardfasern des kalt- und varmbluterherzen. *Pfleugers Arch.* 256: 68-84
10. West, T. C. 1955. Ultramicroelectrode recording from the cardiac pacemaker. *J. Pharmacol. Exp. Ther.* 115:283-90
11. Wenckebach, K. F. 1906. Beitrage zur Kenntnis der menschlichen herztatigkeit. *Arch. Anat. Physiol.* 297:297-354
12. Laslett, E. E. 1908-1909. Syncopal attacks, associated with prolonged arrest of the whole heart. *Q. J. Med.* 2:347-55
13. Eyster, J. A. E., Evans, J. S. 1915. Sinoauricular heart block: with report of a case in man. *Arch. Intern. Med.* 10:832-45
14. Levine, S. A. 1916. Observations on sinoauricular heart block. *Arch. Intern. Med.* 17:153-75
15. Lown, B. 1967. Electrical reversion of cardiac arrhythmias. *Br. Heart J.* 29:469-89
16. Bouvraïn, Y., Slama, R., Temkine, J. 1967. Le block sinoauriculaire et les "maladies du sinus." Reflexions à propos de 63 observations. *Arch. Mal. Coeur Vaiss.* 60:753-73
17. Ferrer, M. I. 1968. The sick sinus syndrome in atrial disease. *J. Am. Med. Assoc.* 206:645-46
18. Tabatznik, B., Mower, M., Sampson, E. B., Prempre, A. 1969. Syncope in the "sluggish sinus node syndrome." *Circulation* 40: Suppl. III p. 200
19. Eraut, D., Shaw, B. D. 1971. Sinus bradycardia. *Br. Heart J.* 33:742-49
20. Silverman, L. F., Mankin, H. T., McGoon, D. C. 1968. Surgical treatment of an inadequate sinus mechanism by implantation of a right atrial pacemaker electrode. *J. Thorac. Cardiovasc. Surg.* 55:264-70
21. Easley, R. M., Goldstein, S. 1971. Sinoatrial syncope. *Am. J. Med.* 50:166-77
22. Short, D. H. 1954. The syndrome of alternating bradycardia and tachycardia. *Br. Heart J.* 16:208-14
23. Rubenstein, J. J., Schulman, C. L., Yurchak, P. M., DeSanctis, R. W. 1972. Clinical spectrum of the sick sinus syndrome. *Circulation* 46:5-13
24. Moss, A. J., Davis, R. J. 1974. Brady-tachy syndrome. *Prog. Cardiovasc. Dis.* 16:439-54
25. James, T. N., Sherf, L., Fine, G., Azorides, R. M. 1966. Comparative ultrastructure of the sinus node in man and dog. *Circulation* 34:139-63
26. Brownlee, W. C., Evans, R. C., Shaw, D. B. 1975. Pathology of conducting system in sino-atrial disease. *Br. Heart J.* 37:779
27. Davis, M., Pomerance, A. 1972. Quantitative study of ageing changes in the human sino-atrial node and internodal tracts. *Br. Heart J.* 34:150-52
28. Shaw, D. B. 1976. The etiology of sino-atrial disorder (Sick Sinus Syndrome). *Am. Heart J.* 92:539-40
29. Strauss, H. C., Bigger, J. T. Jr. 1972. Electrophysiological properties of the rabbit perinodal fibers. *Circ. Res.* 31:490-506
30. Strauss, H. C., Bigger, J. T. Jr., Bassett, A. L., Hoffman, B. F. 1968. The action of diphenylhydantoin on the electrical properties of isolated rabbit and canine atria. *Circ. Res.* 23:463-77
31. Strauss, H. C., Bigger, J. T. Jr. 1971. Slowed conduction, block and reentry in perinodal fibers. *Circulation* 44: Suppl. II, p. 74
32. James, T. N. 1961. Myocardial infarction and atrial arrhythmias. *Circulation* 24:761-76
33. James, T. N. 1968. The coronary circulation and conduction system in acute myocardial infarction. *Prog. Cardiovasc. Dis.* 10:410-49
34. Rosen, K. M., Loeb, H. S., Sinno, M. Z., Rahimtoola, S. H., Gunnar, R. M. 1971. Cardiac conduction in patients with symptomatic sinus node disease. *Circulation* 43:836-44
35. Engel, T. R., Bond, R. C., Schaal, S. F. 1976. First degree sinoatrial heart block. *Am. Heart J.* 91:303-10
36. Goldberg, A. D., Raftery, E. B., Cashman, P. M. M. 1975. Ambulatory electrocardiographic records in patients with transient cerebral attacks or palpitation. *Br. Med. J.* 4:569-71
37. Lipski, J., Cohen, L., Espinoza, J., Motro, M., Dack, S., Donoso, E. 1976. Value of Holter monitoring in assessing cardiac arrhythmias in symptomatic patients. *Am. J. Cardiol.* 37:102-7
38. Reiffel, J. A., Bigger, J. T. Jr., Cramer, M., Reid, D. S. 1977. Ability of Holter electrocardiographic recording and atrial stimulation to detect sinus node dysfunction in symptomatic and asymptomatic patients with sinus bradycardia. *Am. J. Cardiol.* 40:189-94
39. Crook, B. R. M., Cashman, P. M. M., Stott, F. D., Raftery, E. B. 1973. Tape monitoring of electrocardiogram in ambulant patients with sinoatrial disease. *Br. Heart J.* 35:1009-13

40. Bleifer, S. B., Bleifer, D. J., Haussmann, D. R. et al. 1974. Diagnosis of occult arrhythmias by Holter electrocardiography. *Prog. Cardiovasc. Dis.* 16:569-99
41. Hinkle, L. E. Jr., Carver, S. T., Stevens, M. 1969. The frequency of asymptomatic disturbances of cardiac rhythm and conduction in middle-aged men. *Am. J. Cardiol.* 24:629-50
42. Gilson, J. S., Holter, N. J., Glasscock, W. R. 1964. Clinical observations using electrocardiocorder-AVSEP continuous electrocardiographic system. Tentative standards and typical patterns. *Am. J. Cardiol.* 14:204-17
43. Brodsky, M., Wu, D., Denes, P., Kanakis, C., Rosen, K. M. 1977. Arrhythmias documented by 24 hour continuous electrocardiographic monitoring in 50 male medical students without apparent heart disease. *Am. J. Cardiol.*
44. Ferrer, M. I. 1973. The sick sinus syndrome. *Circulation* 47:635-41
45. Fairfax, A. J., Lambert, C. D., Leatham, A. 1976. Systemic embolism in chronic sinoatrial disorder. *N. Engl. J. Med.* 295:190-2
46. Gaskell, W. M. 1884. On the innervation of the heart with special reference to the tortoise. *J. Physiol.* 4:43-127
47. Vassalle, M. 1977. The relationship among cardiac pacemakers: Overdrive suppression. *Circ. Res.* 41:269-77
48. Mandel, W., Hayakawa, H., Danzig, R., Marcus, H. S. 1971. Evaluation of sinoatrial node function in man by overdrive suppression. *Circulation* 44:59-66
49. Mandel, W. J., Hayakawa, H., Hirokazu, A., Allen, H. N., Danzig, R., Kermaier, A. I. 1972. Assessment of sinus node function in patients with the sick sinus syndrome. *Circulation* 46:761-69
50. Narula, O. S., Samet, P., Janvier, R. 1972. Significance of the sinus node recovery time. *Circulation* 45:140-58
51. Bigger, J. T. Jr., Strauss, H. C. 1973. The evaluation of sinoatrial function in man. *Med. Coll. Va. Q.* 9:77-85
52. Gupta, P. K., Lichstein, E., Chadda, K. D., Badui, E. 1974. Appraisal of sinus nodal recovery time in patients with sick sinus syndrome. *Am. J. Cardiol.* 34:265-70
53. Chadda, K. D., Banka, V. S., Bodenheimer, M. M., Helfant, R. H. 1975. Corrected sinus node recovery time. *Circulation* 51:797-801
54. Steinbeck, A., Luderitz, B. 1975. Comparative study of sinoatrial conduction time and sinus node recovery time. *Br. Heart J.* 537:956-62
55. Delius, W., Wirtzfeld, A. 1976. Significance of sinus node recovery time in sick sinus syndrome. In *Cardiac Pacing*, ed. B. Luderitz, pp. 25-32. Berlin: Springer
56. Strauss, H. C., Bigger, J. T. Jr., Saroff, A. L., Giardina, E. G. V. 1976. Electrophysiological evaluation of sinus node function in patients with sinus node dysfunction. *Circulation* 53:763-77
57. Bonke, F. I. M. 1976. The sinoatrial node function. See Ref. 55, pp. 5-12
58. Reiffel, J. A., Bigger, J. T. Jr., Cramer, M., Reid, D. S. 1977. Paced cycle lengths of maximal corrected recovery times. *Circulation* 56: Suppl. III, p. 198
59. Bashour, T., Hemb, R., Wickramsekaran, R. 1973. An unusual effect of atropine on overdrive suppression. *Circulation* 49:911-12
60. Strauss, H. C., Bigger, J. T. Jr. 1973. An unusual effect of atropine on overdrive suppression. *Circulation* 48:911
61. Reiffel, J. A., Bigger, J. T. Jr., Giardina, E. G. V. 1975. "Paradoxical" prolongation of sinus nodal recovery time after atropine in the sick sinus syndrome. *Am. J. Cardiol.* 36:98-104
62. Benditt, D. G., Strauss, H. C., Scheinman, M. M., Behar, V. S., Wallace, A. G. 1976. Analysis of secondary pauses following termination of rapid atrial pacing in man. *Circulation* 54:36-41
63. Strauss, H. C., Saroff, A., Bigger, J. T. Jr., Giardina, E. G. V. 1973. Premature atrial stimulation as a key to the understanding of sinoatrial conduction in man. *Circulation* 47:86-93
64. Breithardt, G., Seipel, L. 1976. Effect of premature atrial depolarization on sinus node automaticity in man. *Circulation* 53:920-25
65. Masini, G., Dianda, R., Graziina, A. 1975. Analysis of sino-atrial conduction in man using premature atrial stimulation. *Cardiovasc. Res.* 9:498-508
66. Dhingra, R. C., Amat-y-Leon, F., Wyndham, C., Deedwania, A. C., Wu, D., Denes, P., Rosen, K. M. 1977. Clinical significance of prolonged sinoatrial conduction time. *Circulation* 55:8-15
67. Jordon, J., Yamaguchi, I., Mandel, W. J. 1977. Characteristics of sinoatrial conduction in patients with coronary artery disease. *Circulation* 55:569-74
68. Reid, D. S., Reiffel, J. A., Bigger, J. T. Jr. 1976. Changes in sinus node automaticity by premature atrial stimulation in man. *6th Eur. Congr. Cardiol., Amsterdam, Abstr. Book* 1:354
69. Miller, H. C., Strauss, H. C. 1974. Measurement of sinoatrial conduction time

- by premature atrial stimulation in the rabbit. *Circ. Res.* 35:935-47
70. Goldreyer, B. N., Damato, A. N. 1971. Sinoatrial node entrance block. *Circulation* 44:789-802
 71. Fleischman, P. 1973. Sinoatrial node entrance block. *Circulation* 47:211-12
 72. Han, J., Malozzi, A. M., Moe, G. K. 1968. Sino-atrial reciprocation in the isolated rabbit heart. *Circ. Res.* 22:355-62
 73. Pauley, K. L., Varghese, J. P., Damato, A. N. 1973. Atrial rhythms in response to an early atrial premature depolarization in man. *Am. Heart J.* 85:323-31
 74. Narula, O. S. 1974. Sinus node re-entry: A mechanism of supraventricular tachycardia (SVT) in man. *Circulation* 50:1114-28
 75. Childers, R. W., Arnsdorf, M. F., de la Fuente, D. J., Gambetta, M., Svenson, R. 1973. Sinus nodal echoes. Clinical case report and canine studies. *Am. J. Cardiol.* 31:220-31
 76. Pauley, K. L., Ruskin, J. N., Damato, A. N. 1975. Sinus and atrioventricular nodal reentrant tachycardia in the same patient. *Am. J. Cardiol.* 36:810-16
 77. Paritzky, Z., Obayashi, K., Mandel, W. J. 1974. Atrial tachycardia secondary to sinoatrial node reentry. *Chest* 66: 526-29
 78. Weisfogel, G. M., Batsford, W. P., Pauley, K. L., Josephson, M. E., Ogunelu, J. B., Ahktar, M., Seides, S. F., Damato, A. N. 1975. Sinus node reentrant tachycardia in man. *Am. Heart J.* 90:295-304
 79. Dhingra, R. C., Wyndham, C., Amat-y-Leon, F., Denes, P., Wu, D., Rosen, K. M. 1975. Sinus nodal responses to atrial extra stimuli in patients without apparent sinus node disease. *Am. J. Cardiol.* 36:445-52
 80. Reiffel, J. A., Bigger, J. T. Jr., Konstam, M. A. 1974. The relationship between sinoatrial conduction time and sinus cycle length during spontaneous sinus arrhythmia in adults. *Circulation* 50: 924-34
 81. Kaplan, B. M., Langendorf, R., Lev, M., Pick, A. 1970. The sick sinus syndrome. *Am. J. Cardiol.* 26:641
 82. Rasmussen, K. 1971. Chronic sinoatrial heart block. *Am. Heart J.* 81:38-47
 83. Wan, S. H., Lee, G. S., Toh, C. C. 1972. The sick sinus syndrome. *Br. Heart J.* 35:942
 84. Engel, T. R., Schaal, S. F. 1973. Digitalis in the sick sinus syndrome: The effects on sinoatrial automaticity and atrioventricular conduction. *Circulation* 48:1201-7
 85. Engel, T. R., Schaal, S. F. 1973. The use of digoxin in the sick sinus syndrome. *Am. J. Cardiol.* 31:129
 86. Bond, R. C., Engel, T. R., Schaal, S. F. 1974. The effect of digitalis on sinoatrial conduction in man. *Am. J. Cardiol.* 33:128 (Abstr.)
 87. Margolis, J. R., Strauss, H. C., Miller, H. C., Gilbert, M., Wallace, A. G. 1975. Digitalis and the sick sinus syndrome. *Circulation* 52:162-69
 88. Dhingra, R. C., Amat-y-Leon, F., Wyndham, C., Wu, D., Denes, P., Rosen, K. M. 1975. The electrophysiological effects of ouabain on sinus node and atria in man. *J. Clin. Invest.* 56:555-62
 89. Bigger, J. T. Jr., Reiffel, J. A., Cramer, M. 1976. Effects of digoxin in patients with sinus bradycardia before and after vagal blockage. *Circulation* 54: Suppl. II, p. 187
 90. Vera, Z., Mokhtar, S., Miller, R. R., Price, G. E., Mason, D. T. 1977. Effect of digoxin on sinus node recovery time and resting and post exercise heart rates in patients with sick sinus syndrome. *Am. J. Cardiol.* 39:264
 91. Cummings, P., McAnulty, J., Rahimtoola, S. 1977. Acute and subacute effects of digitalis in sick sinus syndrome. *Am. J. Cardiol.* 39:290
 92. Ten Eick, R. E., Hoffman, B. F. 1969. Chronotropic effect of cardiac glycosides in cats, dogs, and rabbits. *Circ. Res.* 25:365-78
 93. Schmidt, P. G., Nelson, L. D., Heistad, D. D., Mark, A. L., Abboud, F. M. 1974. Vascular effects of procaine amide in the dog. Predominance of the inhibitory effect on ganglionic transmission. *Circ. Res.* 35:648-960
 94. Mason, J. W., Winkle, R. A., Rider, A. K., Stinson, E. B., Harrison, D. C. 1977. The electrophysiologic effects of quinidine in the transplanted human heart. *J. Clin. Invest.* 59:481-89
 95. Chung, E. K. 1975. Drug associated ECG changes: Quinidine. *Drug Ther.* xx:114-16
 96. Josephson, M. E., Seides, S. F., Batsford, W. P., Weisfogel, G. M., Akhtar, M., Caracta, A. R., Lau, S. H., Damato, A. N. 1974. The electrophysiological effects of intramuscular quinidine on the atrioventricular conducting system in man. *Am. Heart J.* 87:55-64
 97. Hirschfeld, D. S., Ueda, C. T., Roland, M., Scheinman, M. M. 1977. Clinical and electrophysiological effects of intravenous quinidine in man. *Br. Heart J.* 39:309-16

98. Josephson, M. E., Caracta, A. R., Lau, S. H., Gallagher, J. J., Damato, A. N. 1974. Electrophysiologic properties of procainamide in man. *Am. J. Cardiol.* 33:596-603
99. Scheinman, M. M., Weiss, A. N., Shapton, E., Berkowitz, N., Roland, M. 1974. Electrophysiologic effects of procaine amide in patients with intraventricular conduction delay. *Circulation* 49:522-29
100. Calon, A. H., Tiszon, A. R., Akhtar, M., Reddy, C. P., Leeds, C., Damato, A. N. 1976. Quinidine in sinus node dysfunction. *Circulation* 54: Suppl. II, p. 231
101. Wyse, D. G., McAnulty, J. H., Stadius, M., Rahimtoola, S. H. 1978. Electrophysiologic effects of procainamide in patients with normal and prolonged sinoatrial conduction times. *Am. J. Cardiol.* 41:386
102. Befeler, B., Castellanos, A., Wellis, D. E., Vagueiro, M. C., Yeh, B. K. 1975. Electrophysiologic effects of the antiarrhythmic agent disopyramide phosphate. *Am. J. Cardiol.* 35:282-87
103. Josephson, M. E., Caracta, A. R., Lau, S. H., Gallagher, J. J., Damato, A. N. 1973. Electrophysiological evaluation of disopyramide in man. *Am. Heart J.* 86:771-80
104. Strauss, H. C., Scheinman, M. M., Evans, G. T., Beshore, T., Wallace, A. G. 1976. Electrophysiological effects of disopyramide in patients with sinus node dysfunction. *Circulation* 54: Suppl. II, p. 17
105. Yeh, B. K., Sung, P. K., Scherlag, B. J. 1973. Effects of disopyramide on the electrophysiological and mechanical properties of the heart. *J. Pharm. Sci.* 62:1924-29
106. Seides, S. F., Josephson, M. E., Batsford, W. P., Weisfogel, G. M., Lau, S. H., Damato, A. N. 1974. The electrophysiology of propranolol in man. *Am. Heart J.* 88:733-41
107. Wit, A. L., Hoffman, B. F., Rosen, M. R. 1975. Electrophysiology and pharmacology of cardiac arrhythmias. IX. Cardiac electrophysiologic effects of beta adrenergic receptor stimulation and blockade. Part B. *Am. Heart J.* 90:665-75
108. Wit, A. L., Hoffman, B. F., Rosen, M. R. 1975. Electrophysiology and pharmacology of cardiac arrhythmias. IX. Cardiac electrophysiologic effects of beta adrenergic receptor stimulation and blockade. Part C. *Am. Heart J.* 90:795-803
109. Gibson, D., Sowton, E. 1969. The use of beta adrenergic receptor blocking drugs in dysrhythmias. *Prog. Cardiovas. Dis.* 12:16-39
110. Strauss, H. C., Gilbert, M., Svenson, R. H., Miller, H. C., Wallace, A. G. 1976. Electrophysiological effects of propranolol on sinus node function in patients with sinus node dysfunction. *Circulation* 54:452-59
111. Klein, H. O., Jutsin, I., Kaplinsky, E. 1975. Cerebral and cardiac toxicity of a small dose of lidocaine. *Br. Heart J.* 37:775-78
112. Lippestad, C. M., Forfang, K. 1971. Production of sinus arrest by lignocaine. *Br. Med. J.* 1:537
113. Cheng, T. O., Wadhwa, N. 1973. Sinus standstill following intravenous lidocaine administration. *J. Am. Med. Assoc.* 223:790-92
114. Paramiswaran, P., Kahn, D., Monheit, R., Goldberg, H. 1974. Sinus bradycardia due to lidocaine: Clinical-electrophysiologic correlation. *J. Electrocardiol.* 7:75-78
115. Levy, S., Balansard, P., Bouteau, J. M., Gerard, R. 1975. L'arret sinusal: Complication grave de la lidocaine intraveineuse. *Arch. Mal. Coeur Vaiss.* 68:1069-73
116. Wood, R. A. 1971. Sinoatrial arrest: An interaction between phenytoin and lignocaine. *Br. Med. J.* 1:645
117. Jeresaty, R. M., Kahn, A. H., Landry, A. B. Jr. 1972. Sinoatrial arrest in a patient receiving quinidine. *Chest* 61:683-86
118. Agrawal, B. V., Singh, R. B., Vaish, S. K. 1974. Cardiac asystole due to lignocaine with digitalis toxicity. *Acta Cardiol.* 29:341-47
119. Mandel, W. J., Bigger, J. T. Jr. 1971. Electrophysiologic effects of lidocaine on isolated canine and rabbit atrial tissue. *J. Pharmacol. Exp. Ther.* 178:81-93
120. Liberman, N., Harris, R. S., Katz, R. I., Lipschutz, H. M., Dolgin, M., Fisher, V. T. 1968. The effect of lidocaine on the electrical and mechanical activity of the heart. *Am. J. Cardiol.* 22:375-80
121. Dhingra, R. C., Deedwania, P. C., Cummings, J. M., Amat-y-Leon, F., Wu, D., Denes, P., Rosen, K. M. 1978. *Circulation* 57:448-54
122. Unger, A. H., Sklaroff, H. J. 1967. Fatalities following intravenous use of sodium diphenylhydantoin for cardiac arrhythmias. *J. Am. Med. Assoc.* 200:335-36

123. Caracta, A. R., Damato, A. N., Josephson, M. E., Ricciutti, M. A., Gallagher, J. J., Lau, S. H. 1973. Electrophysiologic properties of diphenylhydantoin. *Circulation* 47:1234-41
124. Rosati, R. A., Alexander, J. A., Schaal, S. F., Wallace, A. G. 1967. Influence of diphenylhydantoin on electrophysiological properties of the heart. *Circ. Res.* 21:757-65
125. Bigger, J. T. Jr., Schmidt, D. H., Kutt, H. L. 1968. Relationship between the plasma level of diphenylhydantoin sodium and its cardiac antiarrhythmic effects. *Circulation* 38:363-74
126. Baldessavini, R. J., Lipinski, I. F. 1975. Lithium salts: 1970-1975 *Ann. Intern. Med.* 85:527-33
127. Tilkian, A. G., Schroeder, J. S., Kao, J., Hultgren, H. 1976. The cardiovascular effects of lithium in man: A review of the literature. *Am. J. Med.* 61:665-70
128. Humbert, G., Fillastre, J. P., Leroy, J., Maitrot, B., Tobelem, B., Leroux, G., Lavoine, A. 1974. Intoxication par le lithium. *Sem. Hop.* 50:509-13
129. Tobin, R. J., Nemickas, R., Scanlon, J. P., Moran, J. F., Talano, J. V., Johnson, S., Gunnar, R. M. 1974. ECG of the month. *III. Med. J.* 146:45-46
130. Kleinert, M. 1974. Myokardiopathie unter Lithiumtherapie. *Med. Klin.* 69:494-99
131. Wellens, J. H., Cats, V. M., Duren, D. R. 1975. Symptomatic sinus node abnormalities following lithium carbonate therapy. *Am. J. Med.* 59:285-87
132. Wilson, J. R., Kraus, E. S., Bailas, M. M., Rakita, L. 1975. Reversible sinus node abnormalities following lithium carbonate therapy. *N. Engl. J. Med.* 294:1223-24
133. Habibzadeh, M. A., Zeller, W. H. 1977. Cardiac arrhythmias and hypopotassemia in association with lithium carbonate overdosage. *South. Med. J.* 70:628-30
134. Eliassen, P. 1975. Sinoatrial block during lithium treatment. *Eur. J. Cardiol.* 3:97-98
135. Middelhoff, H. D., Paschen, K. 1974. Lithiumwirkungen auf das EKG. *Pharmakopsychiatr. Neuropsych.* 7:254-64
136. Tilkian, A. G., Schroeder, J. S., Kao, J., Hultgren, H. 1976. Effect of lithium on cardiovascular performance: Report on extended ambulatory monitoring and exercise testing before and during lithium therapy. *Am. J. Cardiol.* 38:701-8
137. Basuray, B. N., Dutta, S. N. 1974. Effects of lithium chloride on cardiovascular responses induced by cholinergic and adrenergic stimulation. *Eur. J. Pharmacol.* 28:387-90
138. Fann, W. E., Davis, J. M., Janowsky, D. S., Cavanaugh, J. H., Kaufman, J. S., Griffith, J. D., Oates, J. A. 1972. Effects of lithium on adrenergic function in man. *Clin. Pharmacol. Ther.* 13:71-76
139. Noma, A., Irisawa, H. 1975. Contribution of an electrogenic sodium pump to the membrane potential in rabbit sinoatrial node cells. *Pfluegers Arch.* 358:289-301
140. Schou, M. 1957. Biology and pharmacology of the lithium ion. *Pharmacol. Rev.* 9:17-58
141. Carmeliet, E. E. 1964. Influence of lithium ions on the transmembrane potential and cation content of cardiac cells. *J. Gen. Physiol.* 47:501-30
142. Ricciutti, M. A., Lisi, K. R., Damato, A. N. 1971. A metabolic basis for the electrophysiologic effects of lithium. *Circulation* 44: Suppl. II, p. 217
143. Furman, S. 1977. Cardiac pacing and pacemakers I: Indications for pacing bradyarrhythmias. *Am. Heart J.* 93:523-30
144. Conde, C. A., Leppo, J., Lipski, J., Stimmel, B., Litwak, R., Donoso, E., Dack, S. 1973. Effectiveness of pacemaker treatment in the bradycardia-tachycardia syndrome. *Am. J. Cardiol.* 32:209-14
145. Chokshi, D. S., Mascarenhas, E., Samet, P., Center, S. 1973. Treatment of sinoatrial rhythm disturbances with permanent cardiac pacing. *Am. J. Cardiol.* 32:215-20
146. Aroesty, J. M., Cohen, S. I., Morkin, E. 1974. Bradycardia-tachycardia syndrome: Results in twenty-eight patients treated by combined pharmacologic therapy and pacemaker implantation. *Chest* 66:257-63, 3:504-7
147. Krishnaswami, V., Geraci, A. R. 1975. Permanent pacing in disorder of sinus node function. *Am. Heart J.* 89:579-85
148. Wohl, A. J., Laborda, N. J., Atkins, J. M., Blomquist, C. G., Mullins, C. B. 1976. Prognosis of patients permanently paced for sick sinus syndrome. *Arch. Intern. Med.* 136:406-8
149. Radford, D. J., Julian, D. G. 1974. Sick sinus syndrome: Experience of a cardiac pacemaker clinic. *Br. Med. J.* 3:504-7
150. Fruenhan, C. T., Heneghan, W. F., Eich, R. H. 1976. Late mortality of patients with sick sinus syndrome. *Circulation* 54: Suppl. II, p. 77

SECRETABLE STORAGE POOLS IN PLATELETS

♦7309

Holm Holmsen, Ph.D.

Specialized Center for Thrombosis Research, Temple University, Philadelphia,
Pennsylvania 19140

Harvey J. Weiss, M.D.

Departments of Medicine, The Roosevelt Hospital and Columbia University
College of Physicians and Surgeons, New York, New York 10019

Knowledge gained during the past two decades has greatly increased our understanding of the roles of platelets in hemostasis, thrombosis, and other biological processes (1-3). However, much remains to be learned. The present review summarizes our current knowledge about granule-bound platelet substances that are released extracellularly after platelet stimulation. We briefly review the *in vitro* responses of platelets to various agonists, emphasizing the granule-bound substances and the mechanism(s) involved in their storage and release. These substances (some of which are also present in plasma, adsorbed on platelets, or present in nonsecretable platelet pools) comprise the secretable storage pool substances of the platelet. Where information is available, we discuss the physiological significance of the secreted substances and their stability following release into the blood. Finally we review congenital and acquired disorders of hemostasis in which the deficiencies of storage pool substances have been described.

TYPICAL PLATELET RESPONSES

Several classes of agents (Table 1) are known to elicit in platelets a sequence of events that has been called the basic platelet reaction (4, 5). Initially, platelets change their shape from discs to spiny spheres in response to low molecular weight substances (except adrenaline) or proteolytic agents, ad-

Table 1 Inducers of human platelet responses (3, 6)

Low molecular weight substances	Proteolytic enzymes	Particulate matter	Agglutinating agents
ADP	Thrombin	Collagen fibrils	Ristocetin + human factor VIII
Epinephrine	Trypsin	Polystyrene latex	Bovine factor VIII Zymosan
Thromboxane A ₂		Glass	Antiplatelet antibody Aggregated immune complexes
Prostaglandins G ₂ and H ₂ ^a	Snake venom proteases	Virus	Polylysine
Vasopressin			
Serotonin			

^aRecent findings indicate that prostaglandins G₂ and H₂ must be converted to thromboxane A₂ in order to stimulate platelets.

here to particulate matter, or agglutinate. (The latter can occur even with metabolically inactive platelets.) Subsequently, metabolically active platelets aggregate, liberate arachidonate from membrane phospholipids and metabolize it to thromboxane A₂, secrete substances stored in dense granules and α -granules, and finally secrete acid hydrolases. It is thought that most of the agonists in Table 1 initially bind to specific receptors on the platelet surface and that a common second messenger is responsible for the subsequent basic platelet reaction (4, 5). The extent to which platelets undergo the entire sequence of events that ends in the secretion of acid hydrolases varies among the different agonists and with the strength of the initial stimulus. It should be stressed that through a series of positive feedback loops (5) and synergistic effects an initially weak stimulation can be followed by a maximal response due to the synthesis and release of the potent platelet aggregant thromboxane A₂ and the secretion of granule-bound ADP.

PLATELET SECRETION

Platelet secretion (6) refers to the specific release of granule-bound substances (with retention of mitochondrial, membranous, and cytoplasmic substances) and should be distinguished from, among others, the release of arachidonate oxidation products, hypoxanthine, lactate, and drug-induced serotonin depletion.

The mechanism of secretion in platelets is not completely understood. The selective release of granule-bound substances is accompanied by the disappearance of granules, is energy dependent, and can be stimulated by calcium ionophores. Therefore, it resembles the secretory process in other cells, but there are some differences. For example, in other cells the secretory products are generally synthesized and packaged in specialized granules, whereas most substances (serotonin is an exception) secreted by platelets are probably synthesized and packaged in the megakaryocyte. In addition, exocytosis has not been demonstrated in platelets, but this difference may be more apparent than real. For example, in platelets the granules move toward the center of the cell and disappear. One secreted substance, fibrinogen, can be demonstrated (after conversion to fibrin) in the surface-connecting canalicular system, which is formed by invaginations of the platelet surface. Hence, platelet granules may fuse with the membranes of the surface-connecting canalicular system and their contents extruded by a process similar to exocytosis.

The secretion of granule-bound substances can be distinguished in several ways. Although the metabolic requirements and the temporal secretion of substances stored in dense granules and α -granules are similar (H. Holmsen and K. Kaplan, unpublished observations; 7), the release of α -granule substances can be achieved with significantly lower concentrations of a release inducer such as thrombin (7). The secretion of acid hydrolases occurs more slowly than the dense-granule- and α -granule-bound substances, requires higher concentrations of inducer, and is more susceptible to the action of metabolic inhibitors. In addition, not all agonists induce the secretion of acid hydrolases.

STORAGE GRANULES IN PLATELETS

Dense Granules

Dense granules (8), approximately 250–300 nm in diameter, are dense in three respects: (a) they may absorb electrons in unfixed, unstained mounts and are therefore electron dense per se; (b) they are highly osmophilic and are the most electron-dense organelles in the platelets after osmium staining; and (c) their specific weight of 1.21 gm/cm³ gives them the highest density of any platelet organelle. Their appearance in the electron microscope is generally spherical, but an elongated form with a round head and a tail is not infrequently seen. In thin sections of platelets fixed and stained by standard methods the dense granules are seen as membrane-enclosed structures containing osmophilic material that only partially fills the lumen. Recent studies (9) indicate, however, that the electron-dense material completely fills the lumen of the organelle in the intact platelet but is

partially lost in preparing the specimens for electron microscopy. With recently developed techniques, relatively pure dense granules, in yields of 20–30%, can be obtained by sucrose gradient centrifugation of platelet homogenates (10).

α -Granules

α -Granules are spherical organelles, 300–500 nm in diameter, and show a great variety of staining characteristics. The ground substance, which always fills the lumen, is either of homogenous electron density or shows an eccentric zone of higher electron density (bull's eye granules). Their specific weight varies between 1.16 and 1.19 gm/cm³. The lightest granules are isolated together with mitochondria after density gradient centrifugation of platelet organelles, while the heavier bull's eye granules are found in fractions of the gradient that are relatively free of mitochondria.

Vesicles

Thin sections of fixed and stained platelets show numerous membrane-enclosed empty structures of various sizes that appear to be cross sections of the surface-connecting channel system (11). Vesicles, 175–250 nm in diameter, containing acid hydrolases and possibly derived from the Golgi apparatus in the megakaryocytes, were recently described in rat platelets (12).

SUBCELLULAR LOCALIZATION OF SECRETABLE STORAGE POOLS

The subcellular localization of granule-bound substances (3, 6) has been inferred from experiments utilizing the techniques of subcellular fractionation, microincineration with x-ray spectroscopy, and cytochemistry, as well as from studies on platelet secretion and on patients with storage pool deficiency (see below). Although no single method is completely satisfactory, the results of these studies have been generally consistent in localizing the various granule-bound substances as shown in Table 2.

CONTENTS OF DENSE GRANULES

The dense granules contain serotonin, calcium, ATP, ADP, and pyrophosphate.

Serotonin

Platelets do not synthesize serotonin but absorb the amine from plasma and store it in the dense granules. Hence, it is possible to label the granule pool of serotonin radioactively by incubating platelets with radiolabeled seroto-

Table 2 Subcellular localization of secretable substances or activities

Dense granules		α -Granules		Vesicles	
Substance or activity	Method ^a	Substance or activity	Method ^a	Substance or activity	Method ^a
Serotonin	A-C, F	Platelet factor 4	A-C	Acid hydrolases	A, C, G
Calcium	A-D	β -Thromboglobulin	A-C		
ATP	A-C, D (?)	Growth factor	A-C		
		Chemotactic factor	A, C		
ADP	A-C, D (?)	Permeability factor	A, C		
		Bactericidal factor	A, C		
Pyrophosphate	A-C, D (?)	Fibrinogen	A, C, E		
		Factor V (?)	A, C		
		Glycosaminoglycan	A, C, F		
		Glycoprotein	A, C		

^aMethod for determination of subcellular localization: A = subcellular fractionation, B = direct analysis of whole platelets from patients with storage pool disorder including those with a decrease of both dense and α -granules, C = secretion, D = microincineration with x-ray spectroscopy, E = direct electron microscopy, F = electron microscopy with radioautography, G = electron microscopy with cytochemical staining.

nin. The release of granule-bound radiolabeled serotonin is currently the method most frequently used to study platelet secretion. It should be noted that serotonin can be released from platelets by nonsecretory processes as well, e.g. by drug-induced depletion of serotonin from storage granules (6).

The mechanism for storing serotonin in the dense granules is not completely clear. One theory holds that serotonin is stored within the granules by complex formation with ATP and a divalent cation (13). Since it has been shown that serotonin is readily incorporated into micelles formed by incubating high concentrations of ATP and metal ions at pH 6.2, the high intragranular concentration of ATP and calcium (Table 3) and the intragranular pH of 6.1 (15) would appear to favor this hypothesis. However, although the deficiency in serotonin parallels that of ATP in platelets from patients with storage pool deficiency (16), it does not correlate with the deficiency in calcium (17). It is therefore possible that serotonin and ATP are stored within the granule by a mechanism that does not require calcium.

Serotonin is released unchanged during secretion, although it is partially oxidized to 5-hydroxytryptophol and 5-hydroxyindole acetate during other release processes (6). Serotonin is relatively stable in plasma and blood, and should be readily available to receptors in the vessel wall. Although it was initially thought that serotonin might be involved in hemostasis, depletion of platelet serotonin by reserpine or treatment with antiserotonin in an intact animal does not increase the bleeding time. At present, the role of serotonin in hemostasis is not clear.

Table 3 Concentrations of constituents within the dense granules^a

Constituent	Amounts secretable ^b ($\mu\text{moles}/10^{11}$ cells)	Concentration in granule (mM) ^c
ATP	2	436
ADP	3	653
Pyrophosphate	1.5	326
Calcium	10	2181
Serotonin	0.3	65

^a The calculations are based on the assumption that the granule is spherical and has an average diameter of 198 nm and that there are 7 dense granules per platelet (49). The volume of each granule is then $6.69 \times 10^{-3} \mu\text{m}^3$ or 0.1% of the platelet volume ($6.8 \mu\text{m}^3$).

^b Values from References 16, 17, 24, 32.

^c Morphometric studies indicate that only 80% of the dense granules undergo exocytosis during maximal secretion (14). This has been accounted for.

Calcium

The calcium concentration of platelets is relatively high. About 60–70% of the platelet calcium is secretable and the theoretical concentration of dense-granule-bound calcium is 2.2 M (Table 3). The subcellular localization of the remaining 30–40% of the platelet calcium is not known. Neither the granular nor the extragranular pool of platelet calcium is exchangeable with extracellular calcium. ATP, ADP, and pyrophosphate can each form tight complexes with calcium and, since the solubility products for these complexes are far exceeded at the concentrations apparently present in the granules (Table 3), it is likely that a substantial fraction of granule-bound calcium exists as insoluble complexes.

The concentration of platelet-derived calcium in blood after complete secretion would be 0.04 mM, whereas the plasma concentration is 2.5 mM. Clearly, secreted platelet calcium contributes little to the plasma concentration, but it could play a role in the interstices of a tightly packed platelet aggregate, which has only a small amount of trapped plasma.

ATP and ADP

Approximately 65% of the total content of the adenine nucleotides in human platelets is stored in the dense granules. The ratio of ATP to ADP in the granules is 2:3 (Table 3). The granule-bound nucleotides do not exchange readily with the adenine nucleotides in cytoplasm. Thus incubation of platelets *in vitro* with radioactive adenosine, adenine, or orthophosphate for 3–4 hr readily labels the cytoplasmic adenine nucleotides (e.g. the metabolic pool, ATP:ADP = 8–10:1), whereas the secretable, granule-bound nucleotides show little radioactivity *in vitro* [although some labeling of ATP has been demonstrated with 10–20 hr incubation (18)]. Incubation of platelets with metabolic inhibitors resulted in depletion of nearly all of

the metabolic pool within two hours but had no effect on the secretable nucleotides (19); this revealed the metabolic inertness of granule-bound storage nucleotides. This finding also indicates that the maintenance of the granule-bound nucleotides in platelets does not depend on active metabolism.

The above findings indicate that granule-bound nucleotides in circulating platelets exchange only slowly with the metabolic pool and they are often referred to as the nonmetabolic pool of platelet adenine nucleotides. In contrast, the storage pool of adenine nucleotides in rabbit platelets is labeled by isotopic precursors *in vivo* (20). However, platelets made devoid of their storage pool of ADP and ATP did not restore these nucleotides after 20 hr in circulation (21). These observations indicate that the storage pool is readily exchangeable at some stage in the production of platelets, presumably in the megakaryocytes where the dense granules are formed.

As indicated above, ATP and ADP in the granules are probably in the form of insoluble calcium complexes. We recently showed (22) that the secretable ATP and ADP (and pyrophosphate) do not give P^{31} -NMR resonance, which provides direct evidence that the storage pool nucleotides are not free in solution; after thrombin-induced secretion the extracellular storage pool did give a P^{31} -NMR resonance that was in good accordance with concurrent chemical measurements. In platelets from patients with storage pool deficiency the deficiency in ADP did not correlate with that of ATP, but it did correlate with the degree of calcium deficiency, which suggests that ATP and ADP may be stored by different mechanisms in the dense granules.

ATP and ADP secreted from platelets rapidly disappear from plasma, with half-lives of 1.5 and 4 min, respectively (23). Most of the ATP and all of the ADP is converted to AMP, which in turn is converted to adenosine, a potent inhibitor of the basic platelet reaction (4). Therefore, the secretory process in platelets results in the extracellular appearance of ADP, which is a platelet agonist, and adenosine, a platelet-inhibiting substance.

Pyrophosphate

The dense-granule constituent pyrophosphate is also metabolically inert (it is not labeled by ^{32}P -orthophosphate *in vitro*) and is not free in solution, as shown by P^{31} -NMR studies. Its storage characteristics and function are not known. Plasma is devoid of inorganic pyrophosphatase and contains very low levels of pyrophosphate (24). *In vitro*, the secretion of the granule-bound pyrophosphate and the formation of pyrophosphate from breakdown of secreted ATP in platelet-rich plasma with normal platelet counts predicts a pyrophosphate level in plasma of 0.02–0.03 mM. One would therefore expect the plasma pyrophosphate levels *in vivo* to show marked increases locally when platelets are activated.

CONTENTS OF THE α -GRANULES

The main constituents of the α -granules (8) appear to be platelet-specific proteins, cationic proteins, and coagulation factors. Secretable platelet proteins were recently extensively reviewed (25). No information about their storage mechanism(s) is available and it is entirely possible that α -granules are heterogenous in their capacities for storing different proteins.

Platelet-Specific Proteins

Platelet-specific proteins are found almost exclusively in platelets and are secreted in response to a variety of agonists. Hence, their presence in the plasma of circulating blood has been regarded as a measure of platelet activation *in vivo*. It should be emphasized that plasma for such studies should be carefully prepared from anticoagulated blood that contains platelet inhibitors to prevent the secretion of these platelet-specific proteins during the collecting and centrifuging procedures.

PLATELET FACTOR 4 Platelet factor 4, a protein composed of a single 7,800-dalton polypeptide chain containing 71 amino acids, readily forms aggregates and is bound to a 59,000-dalton proteoglycan. The platelet factor 4-proteoglycan complex is secreted intact and may be dissociated by high salt concentrations. The proteoglycan may also be displaced by mucopolysaccharides, such as heparin, which form a tight complex with the protein, and much of the activity that has been called "heparin-neutralizing activity" is probably attributable to platelet factor 4, or more specifically high-affinity platelet factor 4 (to distinguish it from other heparin-binding α -granule proteins). The amino acid sequence of high-affinity platelet factor 4 has been determined and it can now be measured in low concentrations by radioimmunoassay.

Since only very small quantities of heparin are present in circulating blood, the significance of the heparin-neutralizing activity of platelet factor 4 is not clear. However, plasma, the vessel wall, and the surface of blood cells might contain proteoglycans that could bind platelet factor 4. Indeed, considering the elegant analytical work that has been done to characterize platelet factor 4, it is surprising how little has been done to characterize its proteoglycan carrier, or to evaluate the binding of platelet factor 4 to other proteoglycans in blood or vessels. Recently, human collagenase was reported to be inhibited by platelet factor 4 (26), but the significance of this effect remains to be clarified.

β -THROMBOGLOBULIN The protein β -thromboglobulin (35,800 dalton) consists of four identical polypeptide chains (each with 81 amino acids of known sequence) and shows great homology with platelet factor 4 (27).

Its heparin-binding activity is distinctly less than that of high-affinity platelet factor 4. Another platelet-specific protein, low-affinity platelet factor 4, has an amino acid sequence probably identical to that of β -thromboglobulin except for an additional, terminal tetrapeptide (S. Niewiarowski, personal communication).

The very short half-life of β -thromboglobulin in the circulation limits its use as an *in vivo* marker for platelet activation. Like high-affinity platelet factor 4, the function of β -thromboglobulin is unknown.

Cationic Proteins

The α -granules probably contain four basic proteins, which, although separated, have not been well characterized chemically. They are currently designated as "factors," and their identifying physiological activities are reflected in their names.

GROWTH FACTOR The growth factor (mitogenic factor) stimulates the growth of certain types of fibroblasts, glial cells, and arterial smooth muscle cells in culture. It probably accounts for the observation that the growth-promoting property of serum prepared from whole blood or platelet-rich plasma is greater than that prepared from platelet-poor plasma. A platelet growth factor that stimulates growth of mouse 3T3 cells has been purified extensively and found to have a molecular weight of 13,700. The growth factor(s) stimulates DNA synthesis and mitotic activity through mechanisms that are not clear at present. Growth factor secreted from platelets adhering to injured blood vessel may stimulate the proliferation of medial smooth muscle cells and hence play an important role in the development of atherosclerosis, but this is not yet proven.

PERMEABILITY FACTOR A cationic 30,000-dalton protein has been partially purified and shown to increase vascular permeability through inducing histamine release from mast cells. This permeability factor is secreted from platelets but its function is unknown.

CHEMOTACTIC FACTOR α -Granules contain an additional cationic protein that activates the fifth component of complement to liberate chemotactic factor from polymorphonuclear leukocytes. However, a lipooxygenase-catalyzed oxidation product of arachidonate, 12-L-hydroxy 5, 8, 10, 14 eicosatetraenoic acid (HETE, which is formed during the basic platelet reaction), also has chemotactic properties (6). At present, the relative contributions of the granule-bound chemotactic factor and HETE to the total chemotactic activity found extracellularly after platelet activation is unclear.

BACTERICIDAL FACTOR Bactericidal factor is classified as a β -lysin, but little is known about its properties or significance.

Coagulation Factors

FIBRINOGEN About 10% of total platelet protein is granule-bound fibrinogen. It is secreted in parallel with platelet factor 4, β -thromboglobulin, and growth factor (7). It does not exchange with extracellular fibrinogen and may be synthesized in the megakaryocyte. In blood the ratio of plasma to platelet fibrinogen is 30:1; hence, platelets do not contribute significant amounts of fibrinogen to the plasma. However, although there are only scanty amounts of plasma between platelets in an aggregate, fibrin strands are readily seen and may well play an important role in transforming the semipermeable platelet aggregate into an impermeable hemostatically effective mass. It is entirely possible that this fibrin is derived from granule-bound platelet fibrinogen, since fibrin strands are frequently seen in the surface-connecting canalicular system.

FACTOR V (Va) The coagulation factor V (Va) (whose exact granule localization is not clear) is present either in an inactive form (V) or becomes activated (Va) when platelets are disrupted or undergo secretion. After secretion, Va is present extracellularly in free solution (28) but may also bind to the platelet membrane. It has been reported that Va is the Xa-binding protein on the platelet surface and that platelet secretion is somehow involved in its appearance (29). These findings may be of great significance since the platelet-Va-Xa complex catalyzed the conversion of prothrombin to thrombin 20 times faster than the phospholipid-Va-Xa complexes.

GLYCOPROTEIN A glycoprotein with molecular weight 147,000–185,000 (reduced form) is secreted from the α -granules (30). Its function is not known.

SUBCELLULAR LOCALIZATION OF ACID HYDROLASES

Human platelets contain acid *p*-nitrophenylphosphatase, β -*N*-acetylglucosaminidase, β -*N*-acetylgalactosaminidase, β -glucuronidase, and β -galactosidase in appreciable amounts, and α -mannosidase, β -glycerophosphatase, arylsulfatase, α -arabinosidase, endoglucosidase, and cathepsins in small amounts (3, 25, 31). Collagenase and proelastase were reportedly present in suspensions of platelets but it is unclear whether contaminating leukocytes might have contributed to this finding (25). All acid hydrolases,

except acid *P*-nitrophenylphosphatase, are secreted from platelets in response to strong stimuli such as thrombin, collagen, and the divalent cationophore A23187. In contrast to other granule-bound substances, which are completely released by strong stimuli, only 30–60% of the total acid hydrolase content is secretable (31). Moreover, β -*N*-acetylglucosaminidase is secreted to a greater extent than β -glucuronidase and β -galactosidase (31), and its secretion is less inhibited by metabolic inhibitors. These observations suggest that β -*N*-acetylglucosaminidase may be stored in organelles different from those containing β -glucuronidase and β -galactosidase, or that the three enzymes are stored in the same organelle but vary in the degree of binding.

The current evidence suggesting that the acid hydrolases are not stored in the α -granules (or dense granules) but in another subcellular compartment may be summarized as follows: First of all, the secretory patterns of acid hydrolases are strikingly different from those of substances stored in dense or α -granules (31, 32). Secondly, a patient with storage pool deficiency has been described whose platelets contain normal amounts of acid hydrolases despite markedly diminished amounts of both α -granules and dense-granule substances (33). Thirdly, megakaryocytes from the rat showed absence of arylsulfatase and acid β -glycerophosphatase in the α -granules, whereas the reaction products of these enzymes were demonstrated in vesicles 150–200 nm in diameter (12). Recent studies on subcellular fractionation tend to support the above findings. The acid hydrolases separate on sucrose gradients slightly displaced from the peak activities of α -granule constituents, but an absolute separation of acid hydrolases from light α -granules has not yet been achieved (8).

The functional significance of acid hydrolase secretion from platelets is not clear. Fibrin clots prepared from platelet-rich plasma lysed more rapidly than those derived from platelet-poor plasma (34), and this could be owing to the presence of proteases, glycosidases, and sulfatases secreted from platelets. In addition, glycosidases secreted from platelets could serve to modify some of the coagulation glycoproteins. The contribution by secretion of platelets to plasma levels of acid hydrolases is relatively small (35). Therefore, if platelet-bound acid hydrolases have any function at all it is probably confined to modifying or regulating events associated with platelet secretion.

STORAGE POOL DEFICIENCY SYNDROMES

General Features

Patients with storage pool deficiency syndromes have mild to severe bleeding problems and their platelets are deficient in the granule-bound ADP

that ordinarily participates only slowly in metabolism [e.g. storage pool ADP (2, 16, 36–40)]. Because of the important role that storage pool ADP plays in platelet aggregation responses, the relative deficiency of this nucleotide in the platelets of patients with storage pool deficiency probably accounts for their hemostatic defect. Granule-bound ATP is also decreased in patients with this disorder. As discussed above, the ADP concentration in the dense granules of normal platelets is higher than that of ATP (ATP:ADP = 2:3), whereas in the cytosol the reverse is true (ATP:ADP = 8–10:1). Hence, with a deficiency of the storage pool adenine nucleotides the percent decrease of the total platelet ADP is greater than that of ATP. This accounts for the significant increase in the ATP to ADP ratio in storage pool deficient platelets (16, 36, 37, 39).

The deficiency of storage pool nucleotides can be demonstrated in yet another way. As indicated above, when normal platelets are incubated for relatively short periods with ^3H or ^{14}C adenine, radioactivity is incorporated into the metabolically active pool of adenine nucleotides but not into the storage pool. The specific radioactivity reflects the proportion of nucleotides in these two pools. In storage pool deficient platelets, the uptake of radioactivity into the metabolic pool is normal. Because of the diminution of storage pool nucleotides, the specific radioactivity is markedly increased, another characteristic feature of storage pool deficient platelets (36, 37). Other dense-granule constituents such as serotonin (16, 36–39), calcium (17), and pyrophosphate (24) are also decreased. The number of granules showing the typical electron core characteristic of dense granules is markedly diminished or absent (40, 41).

The above abnormalities in the platelets of patients with storage pool deficiency suggest that the dense-granule organelles are either absent or functionally defective in this disorder. This abnormality can also be demonstrated by comparing the uptake of radioactive serotonin in storage pool deficient platelets with the uptake in normal platelets. With storage pool deficient platelets, the initial uptake of serotonin (reflecting transport across the plasma membrane) is normal, but saturation levels are decreased owing to either an absence or a diminished storage capacity of the dense granules (16, 38, 39). Because of this defect, serotonin that is absorbed by platelets may be less protected from mitochondrial monoamine oxidases. This would explain the observation that the platelets of patients with storage pool deficiency metabolize ^{14}C serotonin more rapidly than do normal platelets (16, 38, 39). The basic nature of the dense-granule defect in patients with congenital storage pool disease remains to be determined. Since dense granules could originate from the Golgi apparatus of megakaryocytes, a defect in granule development in megakaryocytes might account for the defect in this disorder. A platelet defect similar to that of storage pool disease has

been described in an inbred (fawn-hooded) strain of rat. In this animal, the storage pool defect appears to be present in young platelets newly emerged from the megakaryocyte (42).

Heterogeneity of the Storage Pool Deficiency Syndromes

Patients with congenital storage pool deficiency appear to comprise a heterogeneous group having as a common feature a diminution of substances stored in the dense granules. Several subgroups, with distinct clinical, biochemical, and ultrastructural features have been identified. A storage pool defect has been described in two disorders in which oculocutaneous albinism is a prominent feature: the Hermansky-Pudlak syndrome (38–40) in which a ceroid-like substance is found in the bone marrow, and the Chediak-Higashi syndrome in which leukocyte lysosomes are abnormal (43). The defect, inherited as an autosomal dominant trait, has been described in a family whose platelets were small and showed a unique phospholipid abnormality (44). Storage pool deficiency has also been described in patients in whom the inheritance is not clear (2) and in patients with the syndrome of thrombocytopenia and absent radius (TAR) (45). Finally, a patient has been described whose platelets are deficient in both α -granules and dense granules (33).

Other Biochemical Findings

In most patients, the content of heparin-neutralizing activity and platelet factor 4 is normal (46). However, platelets of the patient with diminished platelet α -granules (see above) are deficient in heparin-neutralizing activity and factor 4 as well as β -thromboglobulin and the platelet growth factor (33). The findings in this patient were consistent with those studies that, utilizing subcellular fractionation procedures, identified the α -granules as the storage location of platelet factor 4 and heparin-neutralizing activity. Further, they suggest that β -thromboglobulin and growth factor are also stored in α -granules. Other patients with decreased heparin-neutralizing activity have also been described (47). Platelet acid hydrolases (β -N-acetylglucosaminidase, β -glucuronidase, β -galactosidase, and acid *p*-nitrophenylphosphatase) are normal in storage pool disease [including the α -granule-deficient variant (33)]. These findings support the conclusions of other studies that lysosomal enzymes are stored neither in dense granules nor α -granules.

Other Functional and Biochemical Defects

Evidence from several types of studies suggest that the release mechanism may also be abnormal in some patients with storage pool deficiency. The abnormalities that suggest a defect in the release mechanism include impair-

ment in: (a) the release of heparin-neutralizing activity (even when present in normal amounts) (46) and platelet-bound ^{14}C -serotonin (16); (b) the release of acid hydrolases by nonsaturating concentrations of thrombin (32); and (c) the synthesis of prostaglandins (PGE_2 and $\text{PGF}_2\alpha$) after collagen stimulation (48). Recent studies indicate that the synthesis of malondialdehyde (a marker for prostaglandin synthesis), after both collagen and epinephrine stimulation of platelets, is abnormal in patients with storage pool deficiency, whereas malondialdehyde formation after stimulation with arachidonic acid is generally normal (33). Taken together, the above findings strongly suggest a defect in the activation of phospholipase A_2 , but this will have to be confirmed by more direct measurements of this enzyme.

Acquired Storage Pool Deficiency

A deficiency in storage pool nucleotide has been described in some patients with disseminated intravascular coagulation (49) and in a patient who had a circulating antiplatelet antibody (50). The reason for the storage pool defect in these patients requires clarification, but is probably related to the fact that both thrombin and antiplatelet antibodies can injure platelets and induce platelet secretion. In contrast, the storage pool defect described in patients with acute leukemia (51) may be owing to an intrinsic defect in the cell precursor of the megakaryocyte.

Literature Cited

1. Smith, J. B., Macfarlane, D. E. 1974. In *The Prostaglandins*, ed. P. W. Ramwell, Vol. 2, pp. 293-343. New York: Plenum.
2. Weiss, H. J. 1975. Platelet physiology and abnormalities of platelet function. *N. Engl. J. Med.* 293:531-41, 580-88.
3. Holmsen, H., Salganicoff, L., Fukami, M. H. 1977. In *Haemostasis: Biochemistry, Physiology and Pathology*, ed. D. Ogston, B. Bennet, pp. 239-319. London: Wiley.
4. Holmsen, H. 1976. Classification and possible mechanisms of action of some drugs that inhibit platelet aggregation. *Ser. Haematol.* 8:50-80.
5. Holmsen, H. 1977. Prostaglandin endoperoxide-thromboxane synthesis and dense granule secretion as positive feedback loops in the propagation of platelet responses during "the basic platelet reaction." *Thromb. Haemostasis* 38: 1030-41.
6. Holmsen, H. 1978. Platelet secretion ("Release reaction"). In *Mechanisms of Hemostasis and Thrombosis*, ed. C. H. Mielke, Jr., R. Rodvien, pp. 73-111. Miami: Symposia Specialists.
7. Witte, L. D., Kaplan, R. L., Nossel, H. L., Lages, B. A., Weiss, H. L., Goodman, D. S. 1978. Studies of the release from human platelets of the growth factor for cultured human arterial smooth muscle cells. *Circ. Res.* 42:402-9.
8. Fukami, M. H. 1977. Human platelet storage organelles. A Review. *Thrombosis* 38:963-70.
9. Costa, J. L., Hui, S. W. 1977. *In situ* observation of dense-body release from hydrated human platelets. *Biophys. J.* 19:307-13.
10. Fukami, M. H., Bauer, J. S., Stewart, G. J., Salganicoff, L. 1978. An improved method for the isolation of dense storage granules from human platelets. *J. Cell Biol.* 77:389-99.
11. Behnke, O. 1968. Electron microscopical observations on the surface coating of human blood platelets. *J. Ultrastruct. Res.* 24:51-69.
12. Bentfield, M. E., Bainton, D. T. 1975. Cytochemical localization of lysosomal enzymes in rat megakaryocytes and platelets. *J. Clin. Invest.* 56:1635.
13. Pietscher, A., Da Prada, M., Berneis, K. H., Tranzer, J. P. 1971. New aspects

- on the storage of 5-hydroxy-tryptamine in blood platelets. *Experientia* 27:993-1120
14. Costa, J. L., Detwiler, T. C., Feinman, R. D., Murphy, D. L., Patlak, C. S., Pettigrew, K. K. 1977. Quantitative evaluation of the loss of human platelet dense bodies following stimulation by thrombin or A23187. *Physiol. J.* 264: 297-306
 15. Wilkins, J. A., Salganicoff, L. 1978. Personal communication
 16. Weiss, H. J., Tschopp, T. B., Rogers, T., Brand, H. 1974. Studies of platelet 5-hydroxytryptamine (serotonin) in storage pool disease and albinism. *J. Clin. Invest.* 54:421-32
 17. Lages, B., Scrutton, M. C., Holmsen, H., Day, H. J., Weiss, H. J. 1975. Metal ion contents of gel-filtered platelets from patients with storage pool disease. *Blood* 46:119-30
 18. Reimers, H. J., Packham, M. A., Mustard, J. F. 1977. Labeling of releasable adenine nucleotides of washed human platelets. *Blood* 49:89-99
 19. Holmsen, H., Setkowsky, C. A., Day, H. J. 1974. Effects of antimycin and 2-deoxyglucose on adenine nucleotides in human platelets. *Biochem. J.* 144: 385-96
 20. Reimers, H. J., Mustard, J. F., Packham, M. A. 1975. Transfer of adenine nucleotides between the releasable and nonreleasable compartments of rabbit blood platelets. *J. Cell Biol.* 67:61-71
 21. Reimers, H.-J., Packham, M. A., Kinlough-Rathbone, R. L., Mustard, J. F. 1977. Adenine nucleotides in thrombin-degranulated platelets: Effect of prolonged circulation *in vivo*. *J. Lab. Clin. Med.* 90:490
 22. Ugurbil, K., Shulman, R. G., Holmsen, H., Costa, J. J. 1978. High resolution ³¹P NMR of human platelets. *Biophys. J.* 21:147a
 23. Holmsen, H., Day, H. J. 1971. Adenine nucleotides and platelet function. *Ser. Haematol.* 4:28-58
 24. Silcox, D. C., Jacobelli, S., McCarthy, D. J. 1973. Identification of inorganic pyrophosphate in human platelets and its release on stimulation with thrombin. *J. Clin. Invest.* 52:1595-1600
 25. Niewiarowski, S. 1977. Proteins secreted by the platelet. *Thromb. Haemostasis* 38:924-938
 26. Hiti-Harper, J., Wohl, H., Harper, E. 1978. Platelet factor 4: An inhibitor of collagenase. *Science* 199:991-93
 27. Begg, G. S., Pepper, D. S., Chesterman, C. N., Morgan, F. J. 1978. Complete covalent structure of human β -thromboglobulin. *Biochemistry* 17:1739-44
 28. Østerud, B., Rapaport, S. I., Lavine, K. K. 1977. Factor V activity of platelets: Evidence for an activated factor V molecule and for a platelet activator. *Blood* 49:819
 29. Majerus, P. W., Miletich, J. P., Jackson, C. M. 1978. Properties of the factor Xa receptor site on human platelets. *Fed. Proc.* 37:1749 (Abstr.)
 30. Hagen, I., Olsen, T., Solum, N. O. 1976. Studies on subcellular fractions of human platelets by the lactoperoxidase-iodination technique. *Biochim. Biophys. Acta* 455:214-25
 31. Holmsen, H., Day, H. J. 1970. The selectivity of the thrombin-induced platelet release reaction: Subcellular localization of released and retained substances. *J. Lab. Clin. Med.* 75:840-55
 32. Holmsen, H., Setkowsky, C. A., Lages, B., Day, H. J., Weiss, H. J., Scrutton, M. C. 1975. Content and thrombin-induced release of acid hydrolases in gel filtered platelets from patients with storage pool disease. *Blood* 46:119-30
 33. Weiss, H. J., Lages, B. A., Witte, L. D., Kaplan, K. L., Goodman, D. W. S., Nossel, H. L., Baumgartner, H. R. 1977. Storage pool disease: Evidence for clinical and biochemical heterogeneity. *Thromb. Haemostasis* 38:3 (Abstr.); Weiss, H. J. 1978. Storage pool disease: Evidence for heterogeneity. *17th Congr. Int. Soc. Hematol. Paris*, p. 90 (Abstr.)
 34. Taylor, F. B., Müller-Eberhard, H. J. 1970. Qualitative description of factors involved in the retraction and lysis of dilute whole blood clots and in the aggregation and retraction of platelets. *J. Clin. Invest.* 49:2068-85
 35. Day, H. J., Aug, G., Holmsen, H. 1972. Platelet release reaction occurring during clotting of platelet-rich plasma. *Proc. Soc. Exp. Med. Biol.* 139:717-21
 36. Holmsen, H., Weiss, H. J. 1970. Hereditary defect in the platelet release reaction caused by a deficiency in the storage pool of platelet adenine nucleotides. *Br. J. Haematol.* 19:643-49
 37. Holmsen, H., Weiss, H. J. 1972. Further evidence for a deficient storage pool of adenine nucleotides in platelets from some patients with thrombocytopathia — "storage pool disease." *Blood* 39:197-209
 38. Haridsty, R. M., Mills, D. C. B., Ketsa-Ard, K. 1972. The platelet defect associated with albinism. *Br. J. Haematol.* 23:689-89

39. Pareti, F. I., Day, H. J., Mills, D. C. B. 1974. Nucleotide and serotonin metabolism in platelets with defective secondary aggregation. *Blood* 44:789-800
40. White, J. G., Edson, J. R., Desnick, S. J., Witkop, C. J. 1971. Studies of platelets in a variant of the Hermansky-Pudlak syndrome. *Am. J. Pathol.* 63:319-30
41. Weiss, H. J., Ames, R. P. 1973. Ultrastructural findings in storage pool disease and aspirin-like defects of platelets. *Am. J. Pathol.* 71:447-60
42. Tschopp, T. B., Weiss, H. J. 1974. Decreased ATP, ADP and serotonin in young platelets of fawn-hooded rats with storage pool disease. *Thromb. Diath. Haemorrh.* 32:670-77
43. Boxer, G. J., Holmsen, H., Robkin, C., Bang, N. U., Boxer, L. A., Baehner, R. L. 1977. Abnormal platelet function in Chediak-Higashi syndrome. *Br. J. Haematol.* 35:521-33
44. Safrit, H. F., Weiss, H. J., Phillips, G. B. 1972. The phospholipid and fatty acid composition of platelets in patients with primary defects of platelet function. *Lipids* 7:60-67
45. Day, H. J., Holmsen, H. 1972. Platelet adenine nucleotide "storage pool deficiency" in thrombocytopenic absent radii syndrome. *J. Am. Med. Assoc.* 221:1053-54
46. Weiss, H. J., Rogers, J. 1973. Platelet factor 4 in platelet disorders. Storage location and the requirement of endogenous ADP for its release. *Proc. Soc. Exp. Biol. Med.* 142:30-35
47. Walsh, P. N., Gagnatelli, G. 1974. Platelet antiheparin activity: Storage sites and release mechanism. *Blood* 44:157-68
48. Willis, A. L., Weiss, H. J. 1972. A congenital defect in platelet prostaglandin production associated with impaired hemostasis in storage pool disease. *Prostaglandins* 4:783-94
49. Costa, J. L., Reese, T. S., Murphy, D. L. 1974. Serotonin storage in platelets: Estimation of storage-packet size. *Science* 183:537-38
50. Zahavi, J., Marder, V. J. 1974. Acquired "storage pool disease" of platelets associated with circulating antiplatelet antibodies. *Am. J. Med.* 56:883-89
51. Cowan, D. H., Grahm, R. C., Baunach, D. 1975. Platelet ultrastructure, adenine nucleotide metabolism and release reaction. *J. Clin. Invest.* 56:188-200

CLONAL ORIGIN OF HUMAN TUMORS

♦7310

Philip J. Fialkow, M.D.

Medical Genetics Section, Medical Service, Veterans Administration Hospital;
Departments of Medicine and Genetics, University of Washington, Seattle,
Washington 98195

Despite the frequency of cancer and years of intensive research, the causes of only a few human neoplasms are known. Also poorly understood are the mechanisms that determine whether a tumor will be benign or malignant, the processes that favor or impede transition from precancerous lesions to malignancy, and the factors that promote metastases or tumor recurrences.

One of the main impediments to progress in this area is the necessary proscription against direct investigation of tumor formation in human subjects. Acceptable alternatives are investigations in animals and indirect studies in man. An indirect approach, which is the subject of this review, takes advantage of naturally occurring cell labels to determine whether a tumor had a single or a multiple cell origin, thereby providing an important clue to the initiating event. For example, one would expect a cancer that arises from a rare event like a "spontaneous" somatic mutation to have a clonal (single cell) origin. On the other hand, a multicellular origin would be anticipated if a neoplasm arose primarily as the result of an oncogenic virus that was highly infectious on a cellular level.

This and similar problems can be investigated through study of neoplasms arising in subjects with two or more genetically distinct cell populations—i.e. in those with cellular mosaicism. Figure 1 schematically illustrates a normal tissue as being composed of two types of cells in equal frequency, each of which is labeled (or "marked") as either A or B. A neoplasm that has a clonal origin begins, by definition, in one cell (e.g. in an A cell), and thus all cells in that tumor will have one type (A) as descendants of the one A progenitor cell. If, in contrast, a tumor is found to contain neoplastic cells of both A and B types, it must have had a multicellular origin.

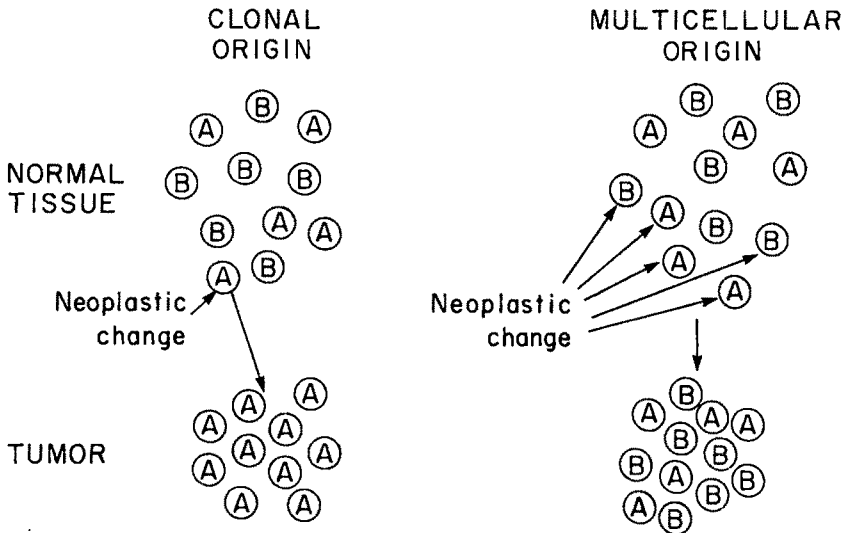


Figure 1 Illustration of the use of cell markers to study tumor development. In this example normal tissue is composed of two types of cells, A and B, in equal frequency. A neoplasm that results from a change occurring in only one cell will have a single marker type (e.g. A), while a tumor that develops from neoplastic transformation of many cells may have both markers, A and B.

Chromosomal mosaicism, in which some cells have an easily identifiable chromosomal abnormality while other cells in the same tissue are normal, has been utilized for this purpose, but its application is limited. Another marker system is dependent upon immunoglobulin (Ig), secreted or on the cell's surface, but it can be applied only to certain lymphoid neoplasms.

A more generally applicable cell marker is dependent upon X-chromosome inactivation and can be used for all tumors that arise in females heterozygous for a gene on the X-chromosome. The X-linked enzyme, glucose-6-phosphate dehydrogenase (G-6-PD) is a particularly useful marker for this purpose. In accordance with X inactivation, only one of the two Xs is active in a given somatic cell; therefore, females with the usual G-6-PD gene (Gd^B) on one X and the common variant Gd^A on the other, have two cell populations, one producing G-6-PD A and the other, G-6-PD B (Figure 2). These enzymes are easily distinguishable by their different electrophoretic mobilities. Tumor cells with a clonal origin have either type B or A G-6-PD, whereas neoplasms with a multicellular origin may have both enzyme types. Since the first application of the G-6-PD system for this purpose by Linder & Gartler (1), many neoplasms have been studied (reviewed in 2). The types of questions that can be answered with G-6-PD markers are well illustrated by investigations of the myeloproliferative disorders.

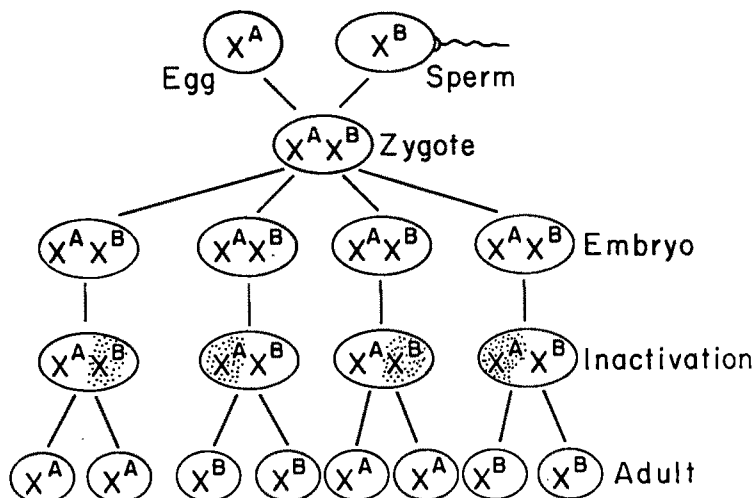


Figure 2 Diagrammatic illustration of X-chromosome inactivation in a female embryo heterozygous for an A and a B gene at the G-6-PD locus. The female zygote inherits one X chromosome from the mother and the other from the father. In this illustration, the maternal X bears an A gene for G-6-PD (X^A) and the paternal X chromosome a B gene (X^B). At some time early in embryogenesis, the two Xs in each somatic cell behave differently—only one is genetically active and the other is inactive. The factors that determine in each cell which X chromosome will be active are unknown, but this is probably a random process. Consequently, on the average, half of the somatic cells have an active X^A and half have an active X^B . Once it is determined which X is to be active in a given cell, that X remains active not only for the lifetime of the cell but also is active in all the cell's progeny. All descendants of a cell with an active X^A will have an active X^A . Thus, the adult female is a mosaic of two cell types—those with an active X^A and those with an active X^B . A tumor with a clonal origin will consist entirely of X^A or X^B cells and, therefore, will contain A or B G-6-PD; a tumor with multicellular origin may contain both X^A and X^B cells and both types of G-6-PD.

MYELOPROLIFERATIVE DISORDERS

Chronic Myelocytic Leukemia

About 85% of patients with chronic myelocytic leukemia (CML) have a very specific and characteristic cytogenetic abnormality, the Philadelphia chromosome (Ph^1) rearrangement. In a typical case, Ph^1 is detected in over 90% of dividing marrow cells and persists throughout the course of the disease.

CLONAL ORIGIN Whereas both B and A enzymes were found in non-hematopoietic tissues from 12 women heterozygous for G-6-PD and with Ph^1 -positive CML, only one type of G-6-PD was observed in their leukemic granulocytes (eight patients typed as B and four as A) (3, 4). Conversely, in Gd^B/Gd^A heterozygotes without blood cell disorders, the granulocytes almost invariably show both B and A enzymes (2). Thus, the finding of

single-enzyme types in CML strongly favors a clonal origin of the disease, a postulate supported by studies with other markers (2). However, the conclusion that CML has a unicellular origin obviously applies only to the stage at which the leukemia is studied. Conceivably, many cells could have been affected at an earlier phase, but when CML is clinically overt, only one clone is evident.

STEM CELL ORIGIN Does CML arise in a cell already committed to differentiate along the granulocyte pathway or in a multipotent stem cell? The latter possibility, first suggested by Ph¹ studies, has been confirmed with G-6-PD markers. The same single-enzyme types found in the granulocytes are also found in erythrocytes, platelets, and monocytes/macrophages; thus the CML clone arises in a stem cell common to all these elements (3). In contrast, marrow fibroblasts do not appear to arise from the CML stem cell (3, 4). Because CML patients' blood lymphocytes stimulated by phytohemagglutinin to divide in culture are reportedly Ph¹-negative, some investigators suggest that lymphocytes do not arise from the CML stem cell. This postulate has been confirmed with G-6-PD markers for some T lymphocytes (5), but other lymphoid cells, possibly B or "null" cells or both, arise from the same progenitors that give rise to granulocytes and other marrow cells in CML (5).

Other Myeloproliferative Disorders

G-6-PD studies in polycythemia vera and idiopathic myeloid metaplasia with myelofibrosis have shown that they too involve pluripotent hematopoietic stem cells and suggest that at the time of diagnosis these disorders probably have clonal origin (6, 7). In contrast, fibroblasts cultured from the marrows of such patients do not arise from the abnormal stem cell clones. Thus, despite the fact that marrow fibrosis is the predominant clinical feature in idiopathic myeloid metaplasia, it may be a secondary event and not part of the abnormal clonal proliferation.

Cause and Pathogenesis

Among the causative factors postulated for CML are "spontaneous" or radiation-induced genetic accidents, viruses, and physiologic defects in homeostasis. The "spontaneous" mutation hypothesis predicts clonal origin, and the genetic marker data are in accord with it. The aberrant homeostasis suggestion implies that the basic abnormality is not intrinsic to the marrow cells but is found in the mechanisms that regulate marrow proliferation and maturation. In its simplest form, this hypothesis predicts multicellular origin, and the genetic marker data make it unlikely to be correct. If caused by a virus or by radiation, CML could originate in one or many cells. For

example, if the virus were only one of several factors necessary for the development of CML, or if the oncogenic changes induced by the virus were rare, a clonal origin would be found. Alternatively, the putative oncogenic virus might have specific affinity for the DNA in the affected regions of the chromosomes involved in Ph¹, in which case the chromosome rearrangement could be induced simultaneously in multiple cells. In any event, the probable clonal origin of CML makes very unlikely any hypothesis of pathogenesis based on continuous recruitment to the tumor of hitherto normal stem cells.

According to some theories of pathogenesis, polycythemia vera and idiopathic myeloid metaplasia with myelofibrosis result from "hyperplasia" of normal stem cells in response to aberrant myeloid-stimulating factors. The genetic marker data do not provide support for these hypotheses and are more compatible with neoplastic origin.

Although each of the myeloproliferative disorders involves a pluripotent hematopoietic stem cell, the diseases differ considerably in their clinical manifestations. It is not known whether these manifestations vary because the basic alterations differ in ways that govern the responses of proliferating cells to regulatory factors, or because different levels of stem cells are involved in each disease. These two mechanisms are not mutually exclusive, so both could be operating. In vitro study of normal stem cells provides some evidence for regulatory defects.

Normal Stem Cells

The question of whether there are any residual normal stem cells is very relevant not only to pathogenesis, but also to therapy. Unless normal stem cells are present, treatment designed to "kill" all aberrant stem cells will, if successful, also kill the patient. The fact that, during remission in CML and polycythemia vera, the single-enzyme G-6-PD phenotypes persist provides no evidence for residual normal stem cells. However, if this putative population contributed less than 5% of the total G-6-PD activity it would not be detected. To increase the sensitivity of the system, hematopoietic colonies grown in semisolid medium were studied. Such colonies from normal G-6-PD heterozygotes have single-enzyme phenotypes and arise from single cells. Thus, analysis of a single colony is equivalent to study of the single progenitor cell from which it was derived. Granulocytic colonies grown from a patient with CML essentially all have the same G-6-PD type as the leukemia clone, providing no evidence for residual normal committed granulocyte stem cells in this disease (4). In contrast, analyses of granulocytic and erythroid colonies from patients with polycythemia vera indicate the presence of stem cells that are not from the clone detected in the blood and are presumably normal residual stem cells (8). However, these cells are

demonstrable *in vitro* only in the presence of erythrocyte- or granulocyte-stimulating factors. Thus, although there are normal stem cells in patients with polycythemia vera, their expression is suppressed *in vivo*. This contrast between polycythemia vera and CML indicates a basic difference in the regulatory abnormalities in the two disorders, although they may involve the same multipotent stem cell.

THE CLONAL ORIGIN OF MOST NEOPLASMS

The causes of human neoplasms are generally unknown, but there is considerable evidence that regardless of the carcinogen, a necessary prerequisite for the development of many if not most human cancers is a mutation-like abnormality in the cells' genetic content (9). Mutation models predict that any given tumor develops from the progeny of one cell. There is little information on the very early stages of tumorigenesis, but there are substantial data on developed neoplasms.

Most benign and malignant tumors studied with cell markers have, like the myeloproliferative disorders, been found to have clonal origin at the time of examination (2). In some instances in which cells with more than one enzyme marker were found in biopsies (e.g. those of some cases of breast carcinoma), one cannot be certain that the specimens studied were homogeneous with respect to cell type. It is difficult to interpret a finding of more than one cell marker in such specimens since admixture with normal cells (e.g. stroma, inflammatory cells, etc) could be responsible for the appearance of the second marker. However, among the tumors for which multicellular origin seems likely are one case of carcinoma of the colon (10), perhaps some invasive carcinomas of the cervix (11, but see also 12), and some spontaneously occurring parathyroid adenomas (13).

Neoplasms of particular interest are those that develop in circumstances in which there is a known tumorigenic factor such as genetic predisposition or environmental factor (e.g. endocrine changes, radiation, viruses, and chemical carcinogens).

HEREDITARY TUMORS

Multiple neurofibromatosis (von Recklinghausen's disease), a disorder inherited as an autosomal dominant, is characterized by skin spots with increased pigmentation and multiple benign tumors consisting predominantly of cells thought to arise from the nerve sheath (neurofibromas). G-6-PD studies indicate that each neurofibroma has multicellular origin, developing perhaps from as many as several thousand cells (14). Origin of

a tumor from so many cells might be seen if a large number of cells were affected simultaneously by the tumorigenic process. Such a mechanism would not be surprising in hereditary neurofibromatosis since it is inherited as an autosomal dominant and, therefore, all of the cells in the target organ are susceptible to neoplasm development by virtue of the inherited mutation. Alternatively, origin of a tumor from many cells could be seen if the oncogenic mechanism initially changed only a single cell and if this alteration subsequently caused neighboring cells to become neoplastic.

The concept of multicellular origin of hereditary neurofibromas cannot be generalized to all hereditary tumors. For example, since the molecular abnormality in xeroderma pigmentosa is defective DNA repair, it is probable that each tumor in that autosomal recessive disease arises from a single cell as the result of a somatic mutation. Similarly, in dominantly inherited retinoblastoma, although every retinoblast has the inherited mutation, Knudson hypothesizes (15) that a retinoblastoma develops only after a cell has acquired a second (somatic) mutation. The prediction from these hypotheses that some inherited tumors in G-6-PD heterozygotes will show only a single enzyme has been confirmed by Baylin et al (16) in the autosomal dominant type II multiple endocrine neoplasia syndrome.

VIRAL TUMORS

The only human neoplasms with proven viral cause are warts. Single-enzyme phenotypes have been observed in six common warts (17), but *condyloma acuminata* ("venereal" warts), rapidly growing infectious "neoplasms" that occur at mucocutaneous junctions such as the vulva and are caused by a papova virus, undoubtedly arise from many cells (18).

Burkitt lymphoma is a malignant disease in man for which there is much circumstantial evidence of a viral cause. The disease occurs occasionally throughout the world, but is notably frequent in areas of Africa and New Guinea. Affected patients usually have multiple lymphoblastic tumors, but studies using G-6-PD and Ig markers suggest strongly not only that each individual growth is a clone, but that all tumors in a given patient are derived from one clone at the time of diagnosis (19). Thus, all tumors in any one patient have the same Ig and single-enzyme G-6-PD types. This fact does not exclude a viral cause for Burkitt lymphoma, but it does suggest that the oncogenic change induced by the putative virus is rare, or that it is only one of several steps necessary for tumorigenesis, or that many cells are altered by the virus initially but only one clone evolves to cause the disease. No matter what the mechanism, clonal origin of the malignant putative viral disease, Burkitt lymphoma, contrasts with the multicellular origin of the benign viral growth, "venereal" wart. This difference suggests

that virus-infected cells give rise to a malignant clone only if one or more additional events occurs.

Most patients with Burkitt lymphoma have therapeutically induced remissions, but the disease reappears in the majority of cases. The G-6-PD and Ig markers in early recurrences (< 3 months) were concordant with those found in the tumors on initial presentation (19), which indicates that early relapses are outgrowths of the original tumors. In contrast, the markers in some late recurrences (> 5 months) were discordant with those found in the original tumors, which indicates that some late relapses may be second primaries (19).

Several growths not proved to be neoplastic have also been studied with G-6-PD and the results have led to formation of stimulating hypotheses. For example, the finding of single-enzyme G-6-PD types in atherosclerotic plaques prompted the suggestion that the lesions are neoplasms, perhaps the results of somatic mutations (20).

CONCLUDING REMARKS

Studies of the neoplasms that arise in patients with mosaicism provide important insights into stem cell relationships and the factors involved in the initiation and progression of the tumors. Most studied neoplasms seem to have clonal origins at the time of diagnosis, and this is compatible with somatic mutation theories of tumorigenesis. However, some hereditary tumors, "venereal" warts, sporadic parathyroid adenomas, one reported case of carcinoma of the colon, and perhaps some invasive cancers of the cervix have multicellular origin. These exceptions suggest that somatic mutation theories of oncogenesis are not valid for all neoplasms.

At present, genetic marker approaches to the study of tumor development are essentially limited to neoplasms that either synthesize Ig or arise in G-6-PD heterozygotes. The electrophoretic G-6-PD variants occur rarely except in black subjects, but when other suitable X-linked markers are discovered, many more populations and tumors can be investigated. It is hoped that such studies will lead to better understanding of the factors that initiate and promote neoplasms in man as well as of the factors underlying remissions and exacerbations.

ACKNOWLEDGMENTS

Studies done in the author's laboratory were supported by Grants GM 15253 and CA 16448 from the Institute of General Medical Sciences and the National Cancer Institute, National Institutes of Health, and by the Medical Research Service of the Veterans Administration.

Literature Cited

1. Linder, D., Gartler, S. M. 1965. Glucose-6-phosphate dehydrogenase mosaicism: utilization as a cell marker in the study of leiomyomas. *Science* 150:67-69
2. Fialkow, P. J. 1976. Clonal origin of human tumors. *Biochim. Biophys. Acta* 458:283-321
3. Fialkow, P. J., Jacobson, R. J., Papayannopoulou, T. 1977. Chronic myelocytic leukemia: clonal origin in a stem cell common to the granulocyte, erythrocyte, platelet and monocyte/macrophage. *Am. J. Med.* 63:125-30
4. Fialkow, P. J., Denman, A. M., Singer, J., Jacobson, R. J., Lowenthal, M. N. 1978. Human myeloproliferative disorders: Clonal origin in pluripotent stem cells. In *Differentiation of Normal and Neoplastic Hematopoietic Cells*, ed. B. Clarkson, P. A. Marks, J. E. Till, Vol. 5, pp. 131-44. Cold Spring Harbor Lab., NY
5. Fialkow, P. J., Denman, A. M., Jacobson, R. J., Lowenthal, M. N. 1978. Chronic myelocytic leukemia: origin of some lymphocytes from leukemic stem cells. *J. Clin. Invest.* 62:815-23
6. Adamson, J. W., Fialkow, P. J., Murphy, S., Prechal, J. F., Steinmann, L. 1976. Polycythemia vera: stem-cell and probable clonal origin of the disease. *N. Engl. J. Med.* 295:913-16
7. Jacobson, R. J., Salo, A., Fialkow, P. J. 1978. Agnogenic myeloid metaplasia: a clonal proliferation of hematopoietic stem cells with secondary myelofibrosis. *Blood* 51:189-94
8. Prechal, J. F., Adamson, J. W., Murphy, S., Steinmann, L., Fialkow, P. J. 1978. Polycythemia vera: the in vitro response of normal and abnormal stem cell lines to erythropoietin. *J. Clin. Invest.* 61:1044-47
9. Koprowski, H., ed. 1977. *Neoplastic Transformation: Mechanisms and Consequences*. Berlin: Dahlem Konferenzen. 313 pp.
10. Beutler, E., Collins, Z., Irwin, L. E. 1967. Value of genetic variants of glucose-6-phosphate dehydrogenase in tracing the origin of malignant tumors. *N. Engl. J. Med.* 276:389-91
11. Smith, J. W., Townsend, D. E., Sparkes, R. S. 1971. Genetic variants of glucose-6-phosphate dehydrogenase in the study of carcinoma of the cervix. *Cancer* 28:529-39
12. Park, I.-J., Jones, H. W. Jr. 1968. Glucose-6-phosphate dehydrogenase and the histogenesis of epidermoid carcinoma of the cervix. *Am. J. Obstet. Gynecol.* 102:106-9
13. Fialkow, P. J., Jackson, C. E., Block, M. A., Greenawald, K. A. 1977. Multicellular origin of parathyroid "adenomas." *N. Engl. J. Med.* 297:696-98
14. Fialkow, P. J., Sagebiel, R. W., Gartler, S. M., Rimoim, D. L. 1971. Multiple cell origin of hereditary neurofibromas. *N. Engl. J. Med.* 284:298-300
15. Knudson, A. G. Jr. 1973. Mutation and human cancer. *Adv. Cancer Res.* 17:317-52
16. Baylin, S. B., Hsu, S. H., Gann, D. S. 1978. Inherited medullary thyroid carcinoma: a final monoclonal mutation in one of multiple clones of susceptible cells. *Science* 199:429-31
17. Murray, R. F. Jr., Hobbs, R. F., Payne, B. 1971. Possible clonal origin of common warts (verruca vulgaris). *Nature* 232:51-52
18. Friedman, J. M., Fialkow, P. J. 1976. Viral "tumorigenesis" in man: cell markers in condylomata acuminata. *Int. J. Cancer* 17:57-61
19. Fialkow, P. J., Klein, E., Klein, G., Clifford, P., Singh, S. 1973. Immunoglobulin and G-6-PD as markers of cellular origin in Burkitt lymphoma. *J. Exp. Med.* 138:89-102
20. Benditt, E. P., Benditt, J. M. 1973. Evidence for a monoclonal origin of human atherosclerotic plaques. *Proc. Natl. Acad. Sci. USA* 70:1753-56

THE SYNDROME OF HYPORENINEMIC HYPOALDOSTERONISM

♦7311

James P. Knochel, M.D.

Department of Internal Medicine, VA Hospital, Dallas, Texas 75216

INTRODUCTION

The syndrome of hyporeninemic hypoaldosteronism (SHH) is a common illness characterized by hyperkalemic, hyperchloremic metabolic acidosis (1-6). It is usually associated with mild or moderately advanced renal insufficiency. Thus, hyperkalemia and acidosis are out of proportion to the degree of renal insufficiency. Plasma renin activity and plasma aldosterone are inappropriately low in those situations in which they should be elevated. Such patients have impaired potassium secretion, and therefore hyperkalemia may first appear under conditions of increased potassium intake or after ingestion of a drug, such as spironolactone, which further impairs potassium secretion by the renal tubule. Electrocardiographic disturbances, heart block, and cardiac syncope have been prominent findings in some patients, and hyperkalemic paralysis has also occurred. However, many patients remain asymptomatic.

SHH may occur in patients with glomerulonephritis or nephrosclerosis and indeed, some glomerular insufficiency is the rule. However, it is most commonly associated with those renal diseases that affect the tubules and interstitium (3, 5). These include idiopathic interstitial nephritis, chronic obstructive uropathy (3), renal lithiasis (3), chronic pyelonephritis (5), lead poisoning (7-9), gout (2), and polycystic disease. One third or more of patients with SHH also have diabetes mellitus (2). Overall, SHH tends to occur more commonly in persons of advanced age. In my own experience, when patients with glomerulonephritis or nephrosclerosis demonstrate characteristic laboratory findings of SHH, its existence is often only tran-

sient, whereas those with disease affecting the renal medulla may demonstrate findings of SHH for several years (10).

Characteristic laboratory features may be seen in SHH besides hyperkalemia and hyperchloremic metabolic acidosis. In those patients whose renal failure is more advanced, serum calcium tends to be high for the degree of azotemia and generally tends to be somewhat higher than in patients with comparable degrees of renal insufficiency associated with glomerular disease. In line with the relative hypercalcemia, calcium excretion into the urine is generally higher in such patients (11).

Etiology of Hyperkalemia

Hyperkalemia in SHH is not associated with anatomic evidence of adrenal insufficiency. Recent studies on patients with SHH (12, 13) produced substantial evidence that hyperkalemia results from decreased production of renin and, in turn, aldosterone. Because of decreased aldosterone, potassium secretion by the renal tubule is impaired, which thereby accounts for decreased ability to excrete ingested potassium (2). There could also exist a defect in potassium transport in other aldosterone receptive tissues.

Considerable published evidence suggests that insulin deficiency may result in persistent hyperkalemia (14, 15). It has also been clearly shown that insulin-deficient diabetics without SHH are relatively intolerant to potassium loads and suffer the danger of hyperkalemia should significant amounts of potassium be administered (2, 6). This has also been demonstrated in experimental animals (16, 17). As stated previously, a substantial percentage of patients with SHH have diabetes mellitus. Insulin has been shown to stimulate transport of sodium and in turn facilitate cellular exchange for potassium in a variety of tissues (17). It is well known that patients with advanced renal insufficiency display resistance to the effects of insulin on glucose transport (11). Whether or not patients with SHH display resistance to the effects of insulin on potassium transport has not been examined.

A third cause of hyperkalemia in SHH could be a defect in potassium adaptation (18, 19). The latter term refers to a state in which an individual becomes tolerant to progressively increasing potassium loads. Thus, doses of potassium that otherwise would cause fatal hyperkalemia are not only excreted more rapidly but are also dissipated into body tissues more rapidly. Such a state probably exists in end-stage renal disease and would satisfactorily explain why some patients with severe renal failure are not hyperkalemic despite normal potassium intake. Therefore, it seems possible that at least part of the hyperkalemia in patients with SHH might result from failure of potassium adaptation.

Acidosis could also contribute to hyperkalemia. It has been known for many years that intracellular accumulation of hydrogen ions may lead to

a reciprocal decline of intracellular potassium concentration and accumulation of potassium ions in extracellular fluid. Nevertheless, in SHH correction of acidosis does not always correct hyperkalemia (20).

Etiology of Metabolic Acidosis

Earlier studies on patients with SHH suggested that a decreased renal mass prevented adequate NH_3 production (21). Although this could certainly contribute to acidosis in some patients, it would not account for decreased NH_3 production in those with essentially normal renal function and no anatomic evidence for reduced renal mass.

The ability to form ammonia is of extreme importance to permit excretion of hydrogen ions into the urine as ammonium. An important effect of hyperkalemia is its propensity to reduce ammonia production by the kidney (22–24). In patients with a diminished filtration rate and especially in those with decreased phosphate excretion into the urine, excretion of hydrogen ion becomes dependent upon the ability to produce ammonia in sufficient quantities.

Recent evidence has shown that hyperkalemia per se not only diminishes the renal capacity to produce ammonia, but also correction of hyperkalemia by potassium-binding resins, such as disodium-polystyrenesulfonate, may substantially increase ammonia production (25). There follows increased excretion of titratable acid and alleviation of metabolic acidosis.

Aldosterone deficiency can also contribute to the generation of metabolic acidosis. It should be kept in mind that aldosterone not only promotes exchange of sodium for potassium ions across the renal tubule, but also promotes exchange of sodium ions for hydrogen ions (26). Thus metabolic acidosis occurs in some cases of adrenal insufficiency. Treatment of SHH with mineralocorticoids may correct both hyperkalemia and acidosis.

Proximal tubular dysfunction might play a contributory role in the pathogenesis of metabolic acidosis in patients with SHH. Recent unpublished observations from this laboratory (27) showed evidence of decreased bicarbonate reabsorptive capacity. The bicarbonate leak and the degree of hyperchloremic acidosis were shown to vary directly with the level of sodium chloride intake. This suggests that if intake of salt is sufficient to maintain an adequate circulatory volume, proximal tubular perfusion will proceed at a rate sufficient to unmask a reabsorptive defect for bicarbonate. However, if the intake of salt and water is diminished so as to allow volume contraction to occur, reabsorption of bicarbonate will become complete. This might explain the observation that patients with SHH commonly demonstrate more pronounced hyperkalemic hyperchloremic acidosis if permitted to determine their own level of sodium chloride intake as they do when outside the hospital environment. Thus, it is common to observe hyperchloremic acidosis in outpatients and then have it disappear after several

days in the hospital as a result of decreased salt intake and, presumably, volume contraction.

Finally, hyperparathyroidism may be an underlying factor in the pathogenesis of depressed proximal tubular bicarbonate reabsorption (28, 29). In high concentrations parathyroid hormone may decrease bicarbonate reabsorption in the proximal tubule. Overproduction of parathormone during the course of chronic renal failure could conceivably result in such a defect. The end result would be hyperchloremic metabolic acidosis. In the presence of renal insufficiency it seems possible that such acidosis could at least contribute to hyperkalemia.

HYPOALDOSTERONISM

Most patients with end-stage renal disease tend to produce either normal or even frankly increased quantities of aldosterone (30). Although unproven, the apparent biologic purpose for increased aldosterone production in chronic renal failure is to enhance secretion of potassium into the bowel contents as an adaptation to prevent hyperkalemia (31). In patients with SHH, hypoaldosteronism has been ascribed to a variety of factors. The most prominent among these appears to be inadequate renin production by the kidney (13). Although it seems possible that hyperkalemia per se may effectively dampen renin production (32), such a hypothesis fails to resolve the issue of why hyperkalemia does not stimulate aldosterone production. In normal subjects, hyperkalemia of the same magnitude observed in patients with SHH is a powerful stimulus to aldosterone production (33). In anephric patients, in whom renin is absent, hyperkalemia is the dominant stimulus to aldosterone production (34). Yet, in SHH, this response to hyperkalemia does not occur. Earlier studies on two patients suggested a defect in 18-hydroxylase activity that could conceivably explain the defect in aldosterone production (35). Nevertheless, others have not confirmed this finding (5, 12). Administration of adrenocorticotrophic hormone (ACTH) to these patients results in normal production of glucocorticoids (3) and their metabolites (12). However, basal as well as ACTH or angiotensin II-stimulated levels of aldosterone are below normal (12). Schambelan and his co-workers (12) proposed that the limited or low capacity to produce aldosterone was the consequence of prolonged hyporeninemia. This view was based upon their experience with patients who have primary aldosteronism with suppressed renin activity and who have undergone successful removal of an adrenal adenoma. Postoperatively, their remaining adrenal fails to produce aldosterone normally in response to either angiotensin II or ACTH (36). A prolonged hyporeninemic state could also blunt stimulated aldosterone production in patients with SHH. Oh and his associates

(37) proposed that both hyporeninemia and hypoaldosteronism might be the result of prolonged suppression of renin and aldosterone production by volume expansion. They showed substantial elevations of extracellular volume in four patients before treatment. Following weight loss and volume contraction, levels of both renin and aldosterone gradually returned toward normal in two of the four in whom follow-up studies were possible. This might indicate that long-standing hypervolemia induced a situation of "disuse atrophy" of the renin-aldosterone axis.

A recent study suggests at least in vitro that regulation of aldosterone production by the ambient concentration of potassium is biphasic in nature (38). Thus, a hypokalemic medium tends to suppress aldosterone formation. An increase in potassium concentration stimulates aldosterone production; however, if levels are too high, aldosterone synthesis is suppressed.

It has been suggested that despite hyperkalemia, concentration of potassium ions in the zona glomerulosa of the adrenal cortex could be abnormally low in patients with SHH (39); therefore aldosterone production would be reduced. Although highly speculative, an abnormally low ambient K concentration at the site of aldosterone synthesis could, in some instances, be the result of insulin deficiency. Insulin deficiency could favor the development of decreased potassium concentration inside the cell as a result of suppressed sodium transport. The difficulty with this postulation is the lack of information as to whether insulin promotes sodium and potassium transport in cells of the zona glomerulosa.

Hyporeninism

As suggested previously, hyperkalemia may depress release of renin from the juxtaglomerular (JG) apparatus in the kidney (32). It has also been suggested that either interstitial disease or, at least in patients with diabetic nephropathy, the anatomic distortion resulting from afferent arteriolar hyalinization and destruction of the JG apparatus could conceivably interfere with either renin release or, indeed, its production (40). As a consequence, those factors causing renin release would be insensitive to hypovolemic stimulation or any other influence that would stimulate release of renin into the circulation. Thus, angiotensin II would not be formed and aldosterone production would fall. It has also been postulated that patients with diabetes mellitus who also display hypoaldosteronism and hyperkalemia may produce an abnormal molecular species of renin (41). The molecule appears to be abnormally large and presumably possesses very little renin activity.

Christlieb (40) found decreased responsiveness of plasma renin to stimuli in patients with adult-onset diabetes mellitus. This has been confirmed by Perez and his co-workers (14). This insensitivity appeared to be worse in those who were hyperkalemic and those with postural hypotension. Even

more impressive was the demonstration of substantially less elevation of plasma aldosterone after tilting. This was also more severe in hyperkalemic group. That age may contribute to diminished renin and aldosterone responsiveness has been shown by several investigators. Indeed, Gossain and his associates (42) reported normal renin activity in otherwise uncomplicated young patients with juvenile diabetes. Based upon the foregoing, it would appear that whatever process may be responsible for suppressing renin release, the depressed production of renin in adults with diabetes mellitus may render them more easily suppressible. This would explain the higher incidence of SHH in adult diabetics. Furthermore, the apparent natural decline of both renin and aldosterone with age would explain the higher incidence of SHH in elderly patients with mild or moderate renal insufficiency.

DIFFERENTIAL DIAGNOSIS

Hyperkalemic hyperchloremic metabolic acidosis without frank renal insufficiency has been observed in a small percentage of patients with disseminated lupus erythematosus (43, 44). In contrast to patients with SHH, renin activity is normal and aldosterone levels are elevated. Further differing from SHH, potassium secretion does not increase normally in response to mineralocorticoids, furosemide, acetazolamide, sodium bicarbonate (43, 44), or ammonium chloride (43). It has been postulated that tubular impairment of potassium secretion occurs in such cases as a result of tubular injury and interstitial nephritis (44). One group of investigators found that serum from a patient with this disorder impaired sodium transport in normal erythrocytes (43). This same patient also showed unusually elevated concentrations of sodium in thermal sweat. Although the precise cause of hyperkalemic hyperchloremic acidosis in lupus erythematosus is unclear, it is clear that dietary potassium restriction and administration of potassium-binding resins are useful in its treatment.

A rare, potassium-secretory defect has recently been described in five members of a family with hyperkalemic hyperchloremic acidosis (45). Similar to patients with lupus, interventions that usually increase potassium secretion such as mineralocorticoid, furosemide, and bicarbonate were ineffective. They differed with respect to renin, which was undetectable; aldosterone was normal.

Rare instances of hyperkalemic hyperchloremic acidosis have been reported in young patients without recognizable renal disease who have hypertension, suppressed renin activity, and very low aldosterone levels (46, 47). Since they correct their abnormalities after sodium restriction, it has been assumed that this is due to excessive sodium reabsorption by the renal tubule.

Primary hypoaldosteronism is a well-recognized cause of hyperkalemic hyperchloremic metabolic acidosis. Some earlier reports of this disease appear to be SHH since mild to moderate renal insufficiency was commonly described in these patients (4, 35). True primary hypoaldosteronism is associated with elevated renin activity. Apparently, sufficient cortisol production persists to prevent extracellular volume contraction. Thus, azotemia is not as typical a finding in primary hypoaldosteronism as it is in total adrenal insufficiency. Otherwise uncomplicated hypoaldosteronism has also been reported after heparin therapy (48). Heparin is known to impair aldosterone production. Addison's disease may rarely be associated with hyperchloremic acidosis, and when it is, hyperkalemia usually precedes its development. Plasma renin activity is very high in such cases.

Although biosynthetic defects of aldosterone have been postulated but unconfirmed in some cases of SHH, authentic cases of 18-hydroxylase deficiency have been observed in children (49). These patients may have hyperkalemic hyperchloremic acidosis.

Treatment

Conventional treatment of SHH has consisted of alkali administered as either NaHCO_3 or Shohl's solution. Though treatment with alkali alone results in some improvement, it is seldom sufficient to correct the chemical abnormalities. Potassium-binding resins are also of value (25). As mentioned earlier, resins may not only relieve hyperkalemia per se, but also the resulting alleviation of hyperkalemia may increase ammonia production by the kidney and, in turn, facilitate increased excretion of acid into the urine. Thereby, a potentially vicious cycle of acidosis and hyperkalemia may be corrected.

Administration of fludrocortisone may also be helpful as a means to treat hyperkalemia as well as facilitate secretion of hydrogen ions and thereby relieve acidosis. Of interest, some patients with SHH may require much larger doses of fludrocortisone than would be required to treat simple adrenal insufficiency.

Finally, in patients who cannot tolerate the risk of additional salt and water retention, judicious administration of loop diuretics such as furosemide may be helpful. When furosemide is given in conjunction with a liberal salt intake, the net effect may be a reduction in total body potassium, correction of hyperkalemia, and improvement of acidosis. Some evidence suggests that prolonged treatment by this means may eventually increase plasma renin activity and improve aldosterone production (37). All patients successfully managed by this means have lost weight. Those with hypertension have frequently become normotensive. This supports the theory that prolonged salt retention and volume expansion may have been responsible initially for reducing renin release and aldosterone production (12, 20).

Literature Cited

1. Sebastian, A. 1976. Metabolic acidosis with special reference to the renal acidosis: Genetic and metabolic injury of the kidney. In *The Kidney*, ed. B. M. Brenner, F. C. Rector, Vol 1, pp. 615-60, Philadelphia: Saunders
2. Carroll, H. J., Farber, S. J. 1964. Hyperkalemia and hyperchloremic acidosis in chronic pyelonephritis. *Metabolism* 13:808-17
3. Weidmann, P., Reinhart, R., Maxwell, M. H., Rowe, P., Coburn, J. W., Massry, S. G. 1972. Syndrome of hyporeninemic hypoaldosteronism and hyperkalemia in renal disease. *J. Clin. Endocrinol. Metab.* 36:965-77
4. Vagnucci, A. H. 1969. Selective aldosterone deficiency. *J. Clin. Endocrinol. Metab.* 29:279-89
5. Vagnucci, A. H. 1970. Selective aldosterone deficiency in chronic pyelonephritis. *Nephron* 7:524-37
6. Cruz, I. A., Dillard, M. G., Hosten, A. O. 1974. Selective hypoaldosteronism: A case report and review of literature. *Med. Ann. DC* 43:597-601
7. Sandstead, H. H., Michelakis, A. M., Temple, T. E. 1970. Lead intoxication. Its effect on the renin aldosterone response to sodium deprivation. *Arch. Environ. Health* 20:356-62
8. Morgan, J. M. 1976. Hyperkalemia and acidosis in lead nephropathy. *South. Med. J.* 69:881-86
9. Gonzales, J. J., Werk, E. E., Thrasher, K., Loadholdt, C. B. 1978. The renin aldosterone system and potassium levels in chronic lead intoxication. *Clin. Res.* 26:47A
10. Knochel, J. P., White, M. G. 1973. Role of aldosterone in renal function. *Arch. Intern. Med.* 131:876-84
11. Knochel, J. P., Seldin, D. W. 1976. The pathophysiology of uremia. See Ref. 1, pp. 1448-85
12. Schambelan, M., Stockigt, J. R., Biglieri, E. G. 1972. Isolated hypoaldosteronism in adults. A renin-deficiency syndrome. *N. Engl. J. Med.* 287:573-78
13. Sebastian, A., Schambelan, M., Lindendorf, S., Morris, R. C. Jr. 1977. Amelioration of metabolic acidosis with fludrocortisone therapy in hyporeninemic hypoaldosteronism. *N. Engl. J. Med.* 297:574-83
14. Perez, G. O., Lesprier, L., Jacobi, J., Oster, J. R., Katz, F. H., Vaamonde, C. A., Fishman, L. M. 1977. Hyporeninemia and hypoaldosteronism in diabetes mellitus. *Arch. Intern. Med.* 137:852-55
15. Perez, G. O., Lesprier, L., Knowles, R., Oster, J. R., Vaamonde, C. A. 1977. Potassium homeostasis in chronic diabetes mellitus. *Arch. Intern. Med.* 137:1018-22
16. Santeusano, F., Faloona, G. R., Knochel, J. P., Unger, R. H. 1973. Evidence for a role of endogenous insulin and glucagon in the regulation of potassium homeostasis. *J. Lab. Clin. Med.* 81:809-17
17. Knochel, J. P. 1977. The role of glucoregulatory hormones in potassium homeostasis. *Kidney Int.* 11:443-52
18. Alexander, E. A., Levinsky, N. G. 1968. An extrarenal mechanism of potassium adaptation. *J. Clin. Invest.* 47:740-48
19. Schultze, R. G., Taggart, D. D., Shapiro, H., Pennell, J. P., Caglar, S., Bricker, N. S. 1971. On the adaptation in potassium excretion associated with nephron reduction in the dog. *J. Clin. Invest.* 50:1061-68
20. Perez, G. O., Lesprier, L. E., Oster, J. R., Vaamonde, C. A. 1977. Effect of alterations of sodium intake in patients with hyporeninemic hypoaldosteronism. *Nephron* 18:259-65
21. Lathem, W. 1958. Hyperchloremic acidosis in chronic pyelonephritis. *N. Engl. J. Med.* 258:1031-38
22. Tannen, R. L., McGill, J. 1976. Influence of potassium on renal ammonia production. *Am. J. Physiol.* 231:1178-84
23. Tannen, R. L. 1977. Relationship of renal ammonia production and potassium homeostasis. *Kidney Int.* 11:453-465
24. Hulter, H. N., Ilnicki, L. P., Harbottle, J. A., Sebastian, A. 1977. *Am. J. Physiol.* 232:F136-F146
25. Szyzlan, P., Better, O. S., Chaimowitz, C., Rosler, A. 1975. Role of hyperkalemia in the metabolic acidosis of isolated hypoaldosteronism. 294:361-65
26. Kurtzman, N. A., White, M. G., Rogers, P. W. 1971. Aldosterone deficiency and renal bicarbonate reabsorption. *J. Lab. Clin. Med.* 77:931-40
27. Barcenas, C. G., Knochel, J. P. Unpublished observations.
28. Rodman, J. S., Heinemann, H. O. 1975. Parathyroid hormone and the regulation of acid-base balance. *Am. J. Med. Sci.* 270:481-89
29. Muldowney, F. P., Donohoe, J. F., Carroll, D. V., Powell, D., Freaney, R. 1972. Parathyroid acidosis in uremia. *Q. J. Med.* 41:321-42

30. Cope, C. L., Person, J. 1963. Aldosterone secretion in severe renal failure. *Clin. Sci.* 25:331-36
31. Hayes, C. P. Jr., McLeod, M. E., Robinson, R. R. 1967. An extrarenal mechanism for the maintenance of potassium balance in severe chronic renal failure. *Trans. Assoc. Am. Physicians* 80:207-15
32. Brunner, H. R., Baer, L., Sealey, J. E., Ledingham, G. G., Laragh, J. H. 1970. The influence of potassium administration and of potassium deprivation on plasma renin in normal and hypertensive subjects. *J. Clin. Invest.* 49:2128-38
33. Funder, J. W., Blair-West, J. R., Coghlan, J. P., Denton, D. A., Scoggins, B. A., Wright, R. D. 1969. Effect of plasma $[K^+]$ on the secretion of aldosterone. *Endocrinology* 85:381-89
34. Williams, G. H., Bailey, G. L., Hampers, C. L., Lauler, D. P., Merrill, J. P., Underwood, R. H., Blair-West, J. R., Coghlan, J. P., Denton, D. A., Scoggins, B. A., Wright, R. D. 1973. Studies on the metabolism of aldosterone in chronic renal failure and anephric man. *Kidney Int.* 4:280-89
35. Jacobs, D. R., Posner, J. B. 1964. Isolated an aldosteronism II. The nature of the adrenal cortical enzymatic defect, and the influence of diet and various agents on electrolyte balance. *Metabolism* 13:522-31
36. Biglieri, E. G., Slaton, P. E. Jr., Silen, W. S. 1966. Postoperative studies of adrenal function in primary aldosteronism. *J. Clin. Endocrinol. Metab.* 26:553-58
37. Oh, M. S., Carroll, H. J., Clemmons, J. E., Vagnucci, A. H., Levison, S. P., Whang, S. M. 1974. *Metabolism* 23:1157-66
38. McKenna, T. J., Island, D. P., Nicholson, W. E., Liddle, G. W. 1978. Dual effects of potassium on aldosterone biosynthesis. *Clin. Res.* 26:15A
39. DeFronzo, R. A., Sherwin, R. S., Felig, P., Bia, M. 1977. Nonuremic diabetic hyperkalemia. *Arch. Intern. Med.* 137:842-43
40. Christlieb, A. P. 1976. Renin-angiotensin-aldosterone system in diabetes mellitus. *Diabetes* 25:820-25
41. DeLeiva, A., Christlieb, A. R., Melby, J. C., Graham, C. A. 1973. Big renin and biosynthetic defect of aldosterone in diabetes mellitus. *N. Engl. J. Med.* 295:639-43
42. Gossain, V. V., Ferrara, E. V., Werk, E. E., Srivastava, L., Privitera, P. J., Hansen, I. 1973. Impaired renin responsiveness. *Arch. Intern. Med.* 132:885-88
43. Hadler, N. M., Gill, J. R., Gardner, J. D. 1972. Impaired renal tubular secretion of potassium, elevated sweat sodium chloride concentration and plasma inhibition of erythrocyte sodium outflux as complications of systemic lupus erythematosus. *Arthritis Rheum.* 15:515-23
44. DeFronzo, R. A., Cooke, C., Goldberg, M., Cox, M., Myers, A. R., Agus, Z. S. 1977. Impaired renal tubular potassium secretion in systemic lupus erythematosus. *Ann. Intern. Med.* 86:268-71
45. Brautbar, N., Levi, J., Rosler, A., Leitesdorf, E., Djaldeti, M., Epstein, M., Kleeman, C. R. 1978. Familial hyperkalemia with normal aldosterone levels. *Arch. Intern. Med.* 138:607-10
46. Gordon, R. D., Geddes, R. A., Pawsey, G. K., O'Halloran, M. W. 1970. Hypertension and severe hyperkalemia associated with suppression of renin and aldosterone and completely reversed by dietary sodium restriction. *Aust. Ann. Med.* 4:287-94
47. Arnold, J. E., Healy, J. K. 1969. Hyperkalemia, hypertension and systemic acidosis without renal failure associated with a tubular defect in potassium excretion. *Am. J. Med.* 47:461-72
48. Wilson, I. D., Goetz, F. C. 1964. Selective hypoaldosteronism after prolonged heparin administration. *Am. J. Med.* 36:635-40
49. Ulick, S., Gautier, E., Vetter, K. K., et al. 1964. An aldosterone biosynthetic defect in a salt-losing disorder. *J. Clin. Endocrinol. Metab.* 24:669-72

DIETARY MANAGEMENT OF DIABETES MELLITUS

♦7312

John W. Ensink, M.D. and Edwin L. Bierman, M.D.

Department of Medicine, University of Washington, Seattle, Washington 98195

INTRODUCTION

The proliferation of scientific and lay articles targeted to modifying dietary habits of Western societies reflects the increasing preoccupation with the national self-image of a corpulent, sedentary citizenry. Despite heightened public awareness that over-consumption of foods of dubious nutritional value may be a risk factor for reduced longevity, the array of dietary notions and admonitions espoused in these publications has provoked much bewilderment as to what constitutes proper dietary management. In this welter of conflicting dietary advice is centered the treatment of patients with diabetes mellitus. This disorder of uncertain cause is characterized by chronic hyperglycemia with additional disturbances of carbohydrate and lipid metabolism attributable to a total or relative deficiency of circulating insulin and with a predilection for development of vascular disease (1, 2). With the increasing life expectancy of Western populations, the morbid sequelae and premature mortality of patients afflicted with diabetes have catalyzed efforts to cure or ameliorate the devastating consequences of this disease. Prior to the advent of insulin, a diabetic was dependent solely upon modification of dietary patterns in the treatment of his disorder. Only during the last decade has a rational approach to dietary management of the diabetic been widely disseminated. However, many physicians and dieticians still hold to the tenet that in the mature overweight non-insulin-requiring diabetic subject, carbohydrate as well as caloric restriction is the benchmark of treatment, whereas in the younger insulin-dependent patient, caloric restraint and judicious selection of carbohydrate-containing food is mandated (3). This traditional approach to diabetic nutrition is entrenched in the misguided logic that the deranged pancreatic B cell would be pro-

tected from further insult by curtailed absorption of sugars from ingested foods. Wood & Bierman highlighted the landmarks of medical thinking throughout the centuries that transformed attitudes to dietary policy and culminated in a modern, more flexible and liberalized dietary prescription for the diabetic (4). It must be emphasized, however, that a conviction that any given dietary regimen, per se, will mitigate the diabetic degenerative complications must be tempered by the fact that its beneficial effects have not been proven by large prospective population studies. The formidable array of problems in attaining this goal is lucidly detailed by West (5). With the proviso that all the answers are not in, we have attempted to place in perspective the role of diet in the management of patients with diabetes mellitus and our perception of a rational approach to this therapeutic objective.

GOALS AND DIETARY MANAGEMENT OF DIABETES

Alleviation or Prevention of the Acute Metabolic Derangement of Insulin Insufficiency

The appropriate dietary prescription, by itself or in conjunction with insulin, may relieve or avert the symptoms of pronounced hyperglycemia, which include listlessness, inordinate weight loss, heightened urinary frequency, genitourinary and cutaneous infections, disturbed visual acuity, and occasional acute painful neuropathy. Obviously, dietary maneuvers alone are ineffectual in patients in whom impaired endogenous insulin reserve is below the threshold for reducing basal hepatic glucose production and increasing postprandial glucose utilization by insulin-dependent tissues. In such patients, dietary manipulation is adjunctive and necessitates supplementation by treatment with sulfonylureas or exogenous insulin. In the older overweight individual in whom B cell secretion is inadequate to overcome peripheral insulin resistance, dietary maneuvers directed to weight reduction are likely to enhance insulin effects and ameliorate metabolic derangements.

Achievement and Maintenance of "Ideal" Body Weight

Diabetes is more prevalent in the highly developed countries where food supplies abound and corpulence is commonplace (1, 5). Aside from the cosmetic advantages, the desirability of attaining an "idealized" body weight is based upon demographic evidence of a high prevalence of diabetes in obese individuals and its mitigation by weight reduction (5, 6). Caloric restriction is a painful process for many and sustained weight reduction is seldom achieved by most patients. Nevertheless, in addition to remission of

symptoms of hyperglycemia in older non-insulin-dependent diabetics, the current public preoccupation with body images may be an incentive to stimulate adherence to a dietary prescription.

Reduction in Atherosclerotic Risk Factors

It is well established that the diabetic is more prone to early development of the atherosclerotic cardiovascular lesions that have become the major killer in this disease (1). Of the several risk factors that hasten atherogenesis, hyperlipidemia has been inculcated as the feature associated with deranged metabolism of the diabetic (7). Therefore, dietary maneuvers that might minimize aberrant lipid metabolism fall within the realm of therapeutic goals for these degenerative diseases in the nondiabetic population.

Retardation of Microvascular Complications

The dire consequences of the angiopathy that characteristically affects the microcirculations of the eye, kidney, nervous system, and limbs are not only devastating to the afflicted patient, but frustrating to the physician who is so frequently helpless in coping with the end-stage organ failure. The relationship of diet to this issue is moot at present, but its potential importance is being resurrected in the debate of whether or not "control" of hyperglycemia will abate the ravages of diabetic microangiopathy.

The goals of dietary treatment apply to all diabetics; however, the interplay of a host of complex factors described in more detail below will influence the achievement of one or more of these objectives in any given patient.

ACHIEVEMENT OF AIMS—PAST AND PRESENT EXPERIENCE

Symptom Abatement—Dietary Composition and Caloric Constraint

Over the span of some 2400 years, records indicate that the diabetic was empirically treated with carbohydrate-rich foods, which were easily accessible in grains and cereals. In 1797 this time-honored dietary recipe was abruptly altered in favor of meals high in animal fat and protein with marked reduction in sugar. The enthusiasm for this divergent dietary prescription originated from the notion that restricted carbohydrate might curtail the degree of glycemia, and the monotonous and unappealing diet assured diminished caloric intake (4). Until the modern era, this dietary regimen remained a cornerstone of diabetic therapy and was deemed vital to survival, despite sporadic evidence that high fat diets worsened carbohydrate tolerance.

Early in this century, a number of scientifically documented studies advocated a reversion to a high carbohydrate, low fat, low caloric diet for diabetic treatment (8–11). Rabinowitch noted: "Potential diabetes can be activated and mild diabetes can be made severe by too rigid restriction of carbohydrate." In a classic study, Himsworth in 1935 showed a consistent improvement in glucose tolerance of normal individuals with progressive elevation of the percentage of carbohydrates in their diet (12). Furthermore, demographic evidence pointed out that diabetes is least common in the most primitive or developing populations where carbohydrate consumption is higher (13). Despite emerging information to the contrary, the indoctrination of patients in the use of low carbohydrate, high fat foods remained the primary dietary dictum until the present. In a survey of English practices conducted by Truswell et al (3) in 1975, it was the universal policy of physicians to restrict carbohydrate to between 35 and 40% of caloric intake. This reluctance to prescribe more liberal amounts of carbohydrate seemingly stems from the preinsulin era when (a) sugar intake was thought to lead to the development of ketoacidosis and (b) dietary carbohydrate was assumed to be detrimental to an already failing beta cell population in the pancreatic islet. Nonetheless, the culpability of sugar in this notion of exhausting the pancreas has never been documented. Epidemiological observations relating sugar intake and prevalence of diabetes have by and large been most difficult to interpret because of the multiple uncontrolled dietary and nondietary variables (13, 14). However, in primitive and developing populations, such as Eskimos, Navajo Indians, and rural African blacks, and those populations in transition, such as the Yemenites in Israel, a rise in the prevalence of diabetes is correlated with the emergence of a more sophisticated animal protein diet associated with a decreased carbohydrate intake (13).

These observations of improved glucose tolerance with higher proportions of dietary carbohydrate and subsequent benefit to diabetic subjects languished until they were resurrected and their importance was reaffirmed by several groups three decades later (15–18). Thus, ingestion of meals of mixed composition with carbohydrate totalling 70% of calories results in decreased fasting and postprandial glucose levels in normal and diabetic subjects. Anderson and colleagues noted this phenomenon in normal men on high sucrose intake as well as on high glucose or augmented starch diets (19). In patients with mild diabetes, Brunzell et al found that, in the face of improved glucose homeostasis, fasting but not postprandial insulin levels were diminished (20). In contrast, high carbohydrate diets tended to worsen hyperglycemia in untreated patients. Following the treatment with insulin or oral sulfonylureas, fasting plasma glucose levels were noted to decrease

as in normal subjects (21). Hence, treated diabetic subjects seemingly do not have further impaired glucose tolerance on antecedent, high carbohydrate intake. As an explanation for this adaptation to high carbohydrate and sugar intake, it has been postulated that there is an increased efficiency of peripheral glucose utilization concomitant with enhanced insulin release to the glycemic stimulus. This mechanistic inference gains some validity from *in vitro* models, since enzymes crucial in the metabolism of glucose, such as hepatic glucokinase and ATP phosphotransferase, are induced by high carbohydrate feeding as is adipose tissue hexokinases Type II, which are integral to glucose utilization by fat tissue (22–24). Of importance, the adaptation of some of these key glucoregulatory enzymes is insulin dependent and is blunted with aging (25, 26). These observations may account for the deterioration in glucose tolerance in poorly controlled diabetics on high carbohydrate intake. Furthermore, although not yet systematically evaluated, it is conceivable that high carbohydrate-containing diets may not be as beneficial in the very elderly as they are in the young.

Thus the credo that a high carbohydrate, isocaloric diet will jeopardize diabetic control is not supported by more rigorous scientific evaluation. In 1971, liberalized carbohydrate intake was formally incorporated into the recommendations of the committee on Food and Nutrition of the American Diabetes Association (ADA) (27). The practical outcome has been to liberalize the proportion of carbohydrate to 50–60% of an isocaloric diet, which necessitates a concomitant decrease in fat. Despite sporadic reports cautioning against such an approach, this stance was reiterated in the most recent dietary guidelines of the ADA (28). This policy also advocated the avoidance of simple sugars or concentrated carbohydrates such as glucose and sucrose, since they tend to lead to rapid oscillations in blood glucose levels, and sucrose has been linked to the genesis of obesity presumably as a result of caloric overconsumption (29). The majority of studies in normal man and in diabetic patients confirm higher blood glucose levels with glucose loading than with isocaloric starch intake. However, to date, despite implications that chronic ingestion of sucrose may be deleterious to diabetes, there is no substantive evidence that sucrose has an adverse affect on blood glucose regulation (30, 31). With perturbations in plasma glucose levels as references, the preponderance of evidence suggests that there is no absolute need to restrict the amount of sucrose or fructose ingested versus complex carbohydrates. Although insulin-dependent diabetics might benefit from less pronounced fluctuations in plasma glucose levels with restriction of sucrose and lactose intake, this question has yet to be resolved.

Corpulence has long been recognized as a factor with the strongest relationship to the risk of developing diabetes. The prevalence of obesity in the

American adult population approximates 20%, with variation according to age and racial origin. In general, black females have the greatest frequency and females of either racial group are fatter than their male counterparts (1). Correspondingly, the incidence of diabetes increases from 0.1% in persons below the age of 20 years to about 15% in those above the age of 60 years. Eighty percent of all diabetics are adults and overweight, and diabetes is frequently precipitated by caloric overindulgence. Obesity is accompanied by increases in fat cell volume, basal and glucose-stimulated insulin concentrations, and insulin resistance with a decrease in insulin receptors (32). Serum triglycerides and glucose incorporation into adipose tissue triglyceride also increase. Weight loss reverses all of these abnormalities (32). It is not surprising, therefore, that the cardinal aim in dietary therapy in the diabetic is to reduce adipose mass by restriction of caloric intake. It is to be emphasized that diabetics should avoid food fads and dietary nostrums perpetrated on the public that wishes to be purged of superfluous fat. A calorically restricted diet should be balanced, nutritionally adequate, and directed to ultimate achievement and long-term maintenance of "ideal" body weight. Success of caloric reduction can often be measured by reversal of the symptoms of acute hyperglycemia, which presumably reflects an attenuation of the insulin-resistant state.

During prolonged partial or total caloric restriction, the body increases the relative contribution of its fat for energy expenditure in an attempt to conserve protein and water. Although a low caloric diet consisting entirely of protein has been suggested to be uniquely advantageous in preserving body nitrogen, it has not been unequivocally shown that protein, per se, is more effective than an isocaloric mixture of protein and carbohydrate (33). Furthermore, the side effects of ketogenesis associated with low carbohydrate (<50 g) diets are potentially deleterious to diabetics. Therefore, a general principle has been to recommend to overweight diabetics a diet balanced in carbohydrate, fat, and protein with restriction of caloric intake varying in total equivalents from 500 to 1200 calories per day, dependent upon amount and rate of weight reduction agreed upon by patient and physician. As emphasized by Allbrink & Davidson: "Adjustment of calories to maintain optimum weight may be the most important part of dietary treatment and perhaps the only justified recommendation" (34). Conversely, by this same principle the first priority in the diet of the underweight young insulin-dependent diabetic should be ample calories to ensure attaining "ideal" body weight. Nutrient intake should be balanced to maintain an "idealized" habitus over the years within the framework of normal living. Thus, experience shows that in those patients who follow the dietary strategems outlined above, the clinical manifestations due to the acute metabolic derangements of diabetes will improve.

Dietary Formulation and Atherogenesis

Several prospective surveys amply documented a link in Western cultures between accelerated atherogenesis and diabetes mellitus irrespective of the treatment modalities employed (35–37). The prevalence of cardiovascular-related mortality is two- to fourfold greater in the diabetic than in the nondiabetic populations, and peripheral vascular disease eventuating in gangrene is twenty times more common in diabetic than in nondiabetic subjects (1). Evidence for increased prevalence of cerebral vascular accidents in the diabetic is sparse. Hence, the susceptibility of the diabetic to the ravages of atherosclerotic vascular disease is unchallenged.

As detailed in excellent reviews (38, 39), the pathogenesis of atherosclerosis is a complex process involving an incompletely understood derangement of the function of the intimal (endothelial) and medial (smooth muscle) cells of arteries. It culminates in a complicated raised lesion consisting of a calcified fibrous plaque with varying degrees of necrosis, thrombosis, and ulceration and leads to symptomatic ischemia and/or infarction of cells distal to the lesion. Risk factors specifically linked to atherogenesis in the diabetic include hypercholesterolemia (cholesterol greater than 150 mg/dl) and hypertriglyceridemia (triglyceride greater than 200 mg/dl). The development of atherosclerosis increases in approximate quantitative relationship to the degree of hyperlipidemia. A widely held hypothesis contends that the higher the ambient levels of lipoprotein, the more likely they are to gain entrance to the arterial wall and overwhelm the ability of smooth muscle cells to metabolize them (38). In the diabetic, the defects in lipid transport are attributable to abnormalities in insulin availability, which is translated into excessive lipoprotein production or defective removal. In any given individual, both production and removal effects may operate. Since insulin is one of the essential factors involved in normal plasma-triglyceride-clearing activity of lipoprotein lipase (LPL) in the severe uncontrolled diabetic, defective clearance is due to inordinately low levels of LPL leading to the accumulation of a variety of lipoproteins. On the other end of the spectrum, the obese patient with mild glucose intolerance characterized by basal and postprandial hyperinsulinemia may have augmented hepatic triglyceride production in conjunction with a defect in clearance. Most diabetics with hypertriglyceridemia fall into the broad range of a more subtle insulin deficiency with "removal lipemia" resulting in accumulation of chylomicrons and low density lipoproteins (LDL) or both. Obesity may predispose to accelerated atherosclerosis not only by aggravating the diabetic state but by an independent mechanism causing hypertriglyceridemia.

Whether or not the adverse consequences of accelerated atherogenesis in the diabetic are due to or can be mitigated by dietary modification is

unsettled. The risk for atherosclerotic vascular disease does not appear to be grossly related to the degree of hyperglycemia. Results from the University Group Diabetes Program study suggested that "control" of blood glucose by insulin did not appear to influence mortality from established atherosclerosis during a five-year period of observation (37). In the orient and in Africa, epidemiological studies failed to corroborate a direct association between diabetes and coronary artery disease (39, 41). It is generally accepted that there is a link between the prevalence of atherosclerosis and the ingestion of the animal fat and protein commonplace in the food of affluent Western societies (42). In contrast, diets high in carbohydrate and low in animal fat are consumed by citizens of most of the less-well-to-do nations, and in Japan where high carbohydrate rice diets are traditional. In the United States by the beginning of this century, the average diet had changed to include more animal protein, less-complex carbohydrates, and more total fat. Hence, the high fat, low carbohydrate diet prescribed for the diabetic paralleled the general dietary practices of the nondiabetic population. It has been further contended that diets high in carbohydrate evoked hypertriglyceridemia in normal subjects and, in certain genetically susceptible individuals, elicited overt lipemia (39, 40, 42). Subsequently, it was found that the critical phenomenon is total caloric intake, since reduction of calories in conjunction with an absolute increase in dietary sugar content is associated with a drop in fasting triglyceride levels (43). Moreover, postprandial triglyceride levels are also lower in patients fed a high carbohydrate diet where both glucose and insulin levels tend to be increased. Therefore, the beneficial effects of a carbohydrate-deprived diet seen in the mildly obese diabetic may be attributable solely to caloric restriction. Unfortunately, no adequate prospective clinical trials to test these questions have been forthcoming.

Observations in small groups of subjects with coronary heart disease indicated modest reductions in circulating triglyceride and cholesterol levels. In a study by Stone & Connor (44), insulin-dependent diabetics on 60% carbohydrate intake showed decreases in serum levels of cholesterol and triglyceride. Weinsier et al (45) were unable to show any increase in mean levels of serum triglyceride in diabetics on high carbohydrate intake, whereas opposite findings were noted by Ginsberg and co-workers (46) in a study of normal subjects. Anderson & Ward followed ten diabetic subjects on a high carbohydrate, high fiber diet for up to 15 months with a reduction in fasting plasma glucose, serum cholesterol, and triglyceride values that led to the discontinuation of insulin and sulfonylureas in eight subjects (47). Although fiber content may in itself influence metabolic parameters because of effects on nutrient absorption, the investigators attributed the lower circulating triglyceride levels to the effects of high carbohydrate.

The role of dietary cholesterol in raising circulating cholesterol levels has recently been questioned on several fronts (39, 42). In an epidemiological survey of cardiovascular disease in Tecumseh, Michigan, serum lipid levels were noted to be independent of dietary habit (48) and a study of men fed whole eggs daily for several weeks failed to indicate a significant increase in serum cholesterol levels (49). However, Connor proposed that individuals already have raised cholesterol levels from dietary intake; a direct relationship between serum and dietary cholesterol is uniformly noted at lower intake (50). In all studies heretofore, it seems clear that saturated fat, as compared with unsaturated fats and oils, leads to hypercholesterolemia. Most of the evidence suggests that the level of plasma cholesterol is determined by the dietary intake of cholesterol, amount of saturated fat, and total calories. Diets containing fat high in unsaturated fatty acids tend to reduce cholesterol levels. At this stage limited information precludes dogmatic statements as to the safety and efficacy of modifying dietary fat composition, though recent epidemiological studies implicate fat in the genesis of cancer (51). Despite the lack of unequivocal proof that dietary modification will ameliorate or prevent coronary artery disease, it seems reasonable to advise diabetic patients as to those dietary components most suspect as risk factors in the atherosclerotic process.

Whether or not dietary manipulation will influence the course of microangiopathic manifestations of diabetes is related to the controversy of whether or not optimal glucose "control" in diabetics will modify the evolution of these lesions. The justification for assiduously regulating the blood sugar of diabetics within the present limitation of insulin injection techniques and dietary maneuvers is controversial (52), and a review of the arguments is beyond the scope of this article. It seems likely that most physicians subscribe to a dietary prescription with or without supplementary insulin and possibly oral sulfonylurea therapy, which tends to "normalize" the excessive blood sugar excursions in the diabetic.

Dietary Compliance

Despite the best intentions of the patient and the physician in endorsing the goals and principles of an "ideal" dietary policy for diabetics, the vast experience in the practical implementation of a prescribed regimen has been far from optimal (5). The odds are roughly three to one that the randomly selected American diabetic has been ignoring his diet and has little or no understanding of what a diet is. The reasons for such a discouraging record are multiple, involving physician comprehension and interest as well as patient understanding, attitudes, and persistence in following the constraints of a diet prescription. Many physicians candidly disagree with the "party line" on liberalized carbohydrate composition, and they misinterpret

the necessity for tailoring dietary intake to the emotional, social, and cultural needs of the diabetic. In addition, many physicians lack an understanding of basic elements of nutrition. This is manifested by a cavalier attitude to its importance, and fosters confusion in the patient as to the role of diet in diabetic therapy. Subsequent failures in the compliance of their patients with dietary advice leads to frustration and anger and further reinforces negative attitudes to dietetic management. The rather dismal compliance rate among patients reflects a number of factors. In general, any constraint on living habits is resented, and a dietary prescription may conflict with life patterns and modify social interaction. The proscription of carbohydrate intake advocated by many professionals is frequently associated with lack of palatability, and although restraint of consumption of refined sugar is widespread, seduction of "junk foods" is real and poses a particular dilemma to the adolescent diabetic faced with the major limitations imposed by dietary exchanges. Often the newly diagnosed adult overweight diabetic, after the shock of diagnosis experiences relief of the symptoms with weight reduction, may arbitrarily relax the strict monitoring of nutrient intake. Furthermore, many patients have no fundamental understanding of the nature of their disorder and may be befuddled by the mass of dietary literature handed to them by a busy practitioner without time to explain in simple terms the reasons for reordering eating habits ingrained over a lifetime.

The basic tenets in attempting to optimize the patient's dietary compliance include: (a) a sympathetic physician versed in objectives of long-term diabetes management; (b) a dietician skilled in imaginatively coping with the patient's individual dietary idiosyncrasies and constructing a flexible diet geared to palatability and minimal monotony; (c) an educational program suited to the level of understanding of each patient; and (d) a continuing follow-up review and revision as necessary by doctor, dietician, and patient of the acceptability of the diet.

FORMULATION OF THE DIABETIC DIET

Calories

In constructing the individual diabetic diet, it is essential to estimate the energy requirements of each patient. These are dependent upon age, sex, vocation, physical activity, and deviation from "ideal" body weight. In the very young or adolescent insulin-dependent patient, calories should be assigned in terms of those needed for basal functions, increased physical activity, and growth. The recommended daily dietary allowances for children of various ages are those prescribed in the ADA Meal Planning Manual (28). For young adults, basal calories are calculated equal to 22

times the desirable body weight (kg). Calories for sedentary, moderate, and strenuous activity are related to the desirable body weight in kg times 6.6, 11, and 22, respectively. Additional calories for women during pregnancy or lactation are individualized. In the overweight patient, the estimate for caloric restriction will be predicated on an optimal time span to achieve ideal body weight and may approximate 500–1200 calories per day. In the elderly in whom lean body mass, resting metabolic rate, and physical activity all decline, caloric intake must be reduced in compensation. A suggested regimen for moderately active people is to decrease calories by 5% for each decade between the ages of 40 and 59 years, by 10% between the ages of 60 and 69, and by a further 10% for the age beyond 70 (53).

Protein

The amount of protein required by the average diabetic is similar to that in the normal person. Daily protein intake should comprise 20% of total calories for growing children and 0.8 g/kg desirable body weight for adults. During catabolic situations such as healing and lactation, protein consumption may be increased two- to threefold if necessary. During ketoacidosis in which nitrogen losses soar, additional protein, two- to threefold times basal intake, is recommended. Since dietary protein derived from meat and dairy products contains variable amounts of saturated fat, a fat-free protein source such as textured vegetable protein may be substituted for animal protein. With age, total body protein declines; however, most elderly persons do not usually require more protein per unit body weight than younger people. Conversely, if renal function in the aged deteriorates, dietary protein should be diminished. The recent fad for weight reduction with a low calorie, liquid diet high in protein should be strongly condemned not only because of the reported deaths in nondiabetics but also because of the likelihood of ketosis and compromised “control” of blood glucose in the diabetic.

Carbohydrate

Within the framework of the total caloric intake for weight maintenance, it is not necessary to restrict carbohydrate. Depending upon individual preference, 50–60% of the daily caloric intake should consist of carbohydrate, the form of which is still under debate. In general, complex carbohydrates, which are polymers of glucose commonly found in cereal grains, tubers, and legumes, are hydrolyzed within the gut thereby releasing glucose more gradually. It is generally agreed that the pure glucose found in corn syrups and in many bakery goods, ice creams, canned goods, and dairy products leads to rapid, wide excursions in blood glucose. Simple sugars differ in their effects. Nonetheless, there appears to be no need to restrict

sucrose intake in the mature non-insulin-dependent diabetic; whereas in the younger insulin-dependent patient, more rapid oscillations of plasma glucose levels are likely, and glucose, sucrose, and lactose probably should be limited. Recent interest has centered on fructose, which is a common sugar found in plants, fruits, and berries and constitutes 50% of sucrose. It is sweeter than either sucrose or glucose. Studies in normal and diabetic subjects failed to incriminate it as a cause of hypertriglyceridemia (54).

The use of polysaccharides such as cellulose, pectin, and guar, which are found in fibrous plants such as bran cereals, improves carbohydrate tolerance (25–40 g of fiber per day). The fibers may act by slowing gastric and intestinal transit time or by interfering with glucose transport across the intestinal wall. In addition, they provide a sense of satiety and increase stool bulk. A potential concern is the possibility of decreased absorption of calcium and other minerals (55). The selection of specific carbohydrate-containing foods should be dictated by the individual response to palatability and rates of gastric motility, digestion, and absorption. Because of variability in day-to-day living with perturbations in physical activity and psychological and metabolic stresses, precision in predicting blood glucose levels after feeding is frequently uncertain, and therefore, in insulin-requiring diabetics, considerable fluctuation in postprandial blood glucose levels is not infrequent.

Fat

Liberalized proportions of carbohydrate automatically necessitate reduced fat intake, and the current recommendations approximate 30–35% of non-protein calories. Fats and oils contain mostly tri- and diglyceride with fatty acids of varying carbon-chain lengths and degrees of saturation. Less than 5% of fat is comprised of phospholipid and sterols, notably cholesterol. Since saturated fats and cholesterol are linked to hyperlipoproteinemia, it is the majority opinion, though not the consensus, that foods low in cholesterol with a higher proportion of polyunsaturated fatty acids may be beneficial (39, 42). The combined restriction of dietary cholesterol in the form of egg yolks, kidney, or liver and the increase of oils with high unsaturated fatty acids, i.e. margarine and corn or sunflower oil may diminish mean cholesterol levels by 20–30%.

Spacing of Meals and Snacks

For the insulin-requiring diabetic patient, eating habits must be adjusted to enable the injected insulin to optimally dispose of absorbed glucose without an inordinate excursion of circulating metabolites in relationship to daily activity. It is recommended that each of the three major meals contain 20–40% of total calories and the remainder be eaten in snacks (28). In the non-insulin-dependent diabetic, snacks are usually omitted whereas in insu-

lin-dependent individuals, supplementary feeding is usually taken at bedtime and at appropriate occasions during the day, depending upon physical activity.

Alcohol

Since alcohol contains seven calories per gram, its use in diabetics has frequently been questioned. In principle, ethanol need not be excluded from the diabetic diet as long as it is not abused and it is considered in the daily total calorie consumption. Both in the insulin-requiring diabetic and in the patient taking oral sulfonylureas, the possibility of severe hypoglycemia exists in a setting of alcoholic excess and inadequate caloric intake. Concern that ethanol may be atherogenic by provoking hypertriglyceridemia must be mitigated by the recent findings that it elevates high density lipoprotein levels, which may prove to lessen the risk for atherosclerosis (42).

Sweetener Alternatives

Since palatability of many foods is related to sweetness, patient adherence to a calorie-restricted diet must take into account the inclination of many patients, particularly the young diabetic, to consume sweetened beverages and food. With the banning of cyclamates and the threatened removal of saccharine as nonnutritive sweeteners, there has been an intensive search for potential substitutes for glucose and sucrose in diabetic diets (56). Of the naturally occurring monosaccharides, fructose, sorbitol, and xylitol are the only ones approximately as sweet as sucrose (range 1.6 through 0.6). Because they are calorically equivalent to glucose and sucrose, their unrestricted use is limited as with noncalorogenic sweeteners, and it is debatable whether they offer any advantage.

Ancillary Dietary Measures

There is no need to enrich the diet of the stable diabetic with vitamins, minerals, electrolytes, or trace elements. In situations such as protracted polyuria, "brittle" diabetes, or diabetic ketoacidosis, vitamins and electrolyte supplementation is often beneficial or mandated. Chromium, a trace element essential to several enzymatic functions, occasionally improves glucose tolerance; however, since it is widely distributed in food, a balanced intake should preclude the need to readjust for its supplementation. In patients with chronic renal failure, protein content should be reduced. In insulin-requiring diabetics, it is essential to anticipate the changing caloric needs during exercise as well as the difficulties in controlling the character and quantity of nutrients when dining out or during foreign travel. Diabetics should be strongly encouraged to consult their physicians or dieticians for suggested dietary modification in advance of major changes in their daily habits.

SUMMARY

The recommended approach to dietary management of the diabetic reinforces long-standing knowledge that palatability and flexibility in selection of foods of higher carbohydrate content are most likely to lead to patient acceptance and mitigation of the symptoms of acute metabolic derangements of diabetes. A continuous striving to achieve and maintain "idealized" body weights throughout one's lifetime is the cornerstone of long-term dietary prescription. It is hoped that additional information will be forthcoming to support the belief that this dietary policy will retard the progression of the vascular complications of diabetes.

ACKNOWLEDGMENT

The authors are grateful to Ms. Suzanne Burian for precise secretarial assistance, and they gratefully acknowledge support of USPHS Grants AM-16008, AM-02456, and GCRC RR-37.

Literature Cited

1. Diabetes Data Compiled 1977. 1978. *DHEW Publication (NIH)78-1468*. Washington: GPO
2. *Natl. Inst. Arthritis, Metab. Dig. Dis. and Juvenile Diabetes Found. Workshop, Philadelphia*. 1977. Etiology and pathogenesis of insulin dependent diabetes mellitus. Washington, DC: GPO
3. Truswell, A. S., Thomas, B. J., Brown, A. M. 1975. Survey of dietary policy in management in British diabetic clinics. *Br. Med. J.* 4:7-11
4. Wood, F. C. Jr., Bierman, E. L. 1972. New concepts in diabetic dietetics. *Nutr. Today* May-June, pp. 4-12
5. West, K. M. 1973. Diet therapy of diabetes: an analysis of failure. *Ann. Intern. Med.* 79:425-34
6. Wall, J. R., Pyke, D. A., Oakley, W. G. 1973. Effect of carbohydrate restriction in obese diabetics: relationship of control to weight loss. *Br. Med. J.* 1:577-78
7. Albrink, M. J., Mann, E. B. 1958. Serum triglycerides in health and disease. *Diabetes* 7:194-201
8. Sansum, W. D., Blatherwick, N. R., Bowden, R. 1926. The use of high carbohydrate diets in the treatment of diabetes mellitus. *J. Am. Med. Assoc.* 86:178-81
9. Richardson, R. 1929. High carbohydrate diets in diabetes mellitus. *Am. J. Med. Sci.* 177:426-30
10. Ellis, A. 1934. Increased carbohydrate tolerance in diabetics following hourly administration of glucose and insulin over long periods. *Q. J. Med.* 3:137-53
11. Rabinowitch, I. M. 1935. Experiences with high-carbohydrate, low-calorie diet for treatment of diabetes mellitus. *Can. Med. Assoc. J.* 33:136-44
12. Himsworth, H. P. 1935. The dietetic factor determining the glucose tolerance and the sensitivity to insulin of healthy men. *Clin. Sci.* 2:67-94
13. West, K. M., Kalbfleisch, J. M. 1971. Influence of nutritional factors on prevalence of diabetes. *Diabetes* 20:99-108
14. Walker, A. R. P. 1977. Sugar intake in diabetes mellitus. *S. Afr. Med. J.* 51:842-51
15. Singh, I. 1955. Low-fat diet and therapeutic doses of insulin in diabetes mellitus. *Lancet* 1:422-525
16. Kempner, W., Peschel, R. L., Schlager, C. 1958. Effect of rice diet on diabetes mellitus associated with vascular disease. *Postgrad. Med.* 24:359-71
17. Ernest, I., Linner, I., Svanborg, A. 1965. Carbohydrate-rich, fat-poor diet in diabetes. *Am. J. Med.* 39:594-600
18. Anderson, J. W., Herman, R. H., Zakim, D. 1968. Glucose tolerance and insulin response to prolonged carbohydrate feeding in normal men. *Am. J. Clin. Nutr.* 21:529-39
19. Anderson, J. W., Herman, R. H., Zakim, D. 1973. Effect of high glucose and high sucrose diets on glucose toler-

- ance of normal men. *Am. J. Clin. Nutr.* 26:600-7
20. Brunzell, J. D., Lerner, R. L., Hazzard, W. R., Porte, D. Jr., Bierman, E. L., 1971. Improved glucose tolerance with high-carbohydrate feeding in mild diabetes. *N. Engl. J. Med.* 284:521-24
 21. Brunzell, J. D., Lerner, R. L., Porte, D. Jr., Bierman, E. L. 1974. Effect of a fat-free, high-carbohydrate on diabetic subjects with fasting hyperglycemia. *Diabetes* 23:138-42
 22. Borreback, B., Hultman, E., H:Son Nilsson, L., Roch-Norlund, A. E., Spydevold, O. 1970. Adaptable glucokinase activity of human liver. *Biochem. Med.* 4:469-75
 23. Sharma, C., Manjeshwar, R., Weinhouse, S. 1963. Effects of diet and insulin on glucose-adenosine-triphosphate-phosphotransferases of rat liver. *J. Biol. Chem.* 238:3840-45
 24. Bernstein, R. S., Marshall, M. D., Carney, A. L. 1977. Effects of dietary composition on adipose tissue hexokinase-II and glucose utilization in normal and streptozocin-diabetic rats. *Diabetes* 26:770-79
 25. Moore, R. O. 1968. Effect of age of rats on the response of adipose tissue to insulin and the multiple forms of hexokinase. *J. Gerontol.* 23:45-49
 26. Adelman, R. C. 1970. An age-dependent modification of enzyme regulation. *J. Biol. Chem.* 245:1032-35
 27. Bierman, E. L., Albrink, M. J., Arky, R. A., Connor, W. E., Dayton, N., Spritz, N., Steinberg, D. 1971. Special report: principles of nutrition and dietary recommendations for patients with diabetes mellitus. *Diabetes* 20:633-34
 28. *A Guide for Professionals: The Effective Application of Exchange Lists for Meal Planning.* 1977. New York, Chicago: Am. Diabetes Assoc. 23 pp.
 29. Arky, R. A. 1974. Current views of the American Diabetes Association Committee on Food and Nutrition regarding carbohydrate in the diabetic diet. In *Is the Risk of Becoming Diabetic Affected by Sugar Consumption?* ed. S. S. Hillebrand, pp. 53-55. Bethesda: Int. Sugar Res. Found.
 30. Crapo, P. A., Reaven, G., Olefsky, J. 1976. Plasma glucose in insulin responses to orally administered simple and complex carbohydrates. *Diabetes* 25:741-47
 31. Arvidsson-Lenner, R. 1976. Specially designed sweeteners and food for diabetics—a real need? *Am. J. Clin. Nutr.* 29:716-25
 32. Rabinowitz, D. 1970. Some endocrine and metabolic aspects of obesity. *Ann. Rev. Med.* 21:241-58
 33. Van Itallie, T. B., Yang, M.-V. 1977. Current concepts in nutrition. Diet and weight loss. *N. Engl. J. Med.* 297:1158-61
 34. Albrink, M. J., Davidson, P. C. 1971. Dietary therapy and prophylaxis of vascular disease in diabetics. *Med. Clin. North Am.* 55:877-87
 35. Garcia, M., McNamara, P., Gordon, T., Kannel, W. B. 1974. Morbidity and mortality in diabetics in the Framingham population: 16 year follow-up study. *Diabetes* 23:105-11
 36. Pel, S., D'Alonzo, C. A. 1970. Factors associated with longterm survival of diabetics. *J. Am. Med. Assoc.* 214:1833-40
 37. University Groups Diabetes Program. 1970. A study of the effects of hypoglycemia agents in vascular complications in patients with adult onset diabetes: II. Mortality results. *Diabetes* 19: Suppl. II, pp. 787-830
 38. Ross, R., Glomset, J. A. 1976. The pathogenesis of atherosclerosis. *N. Engl. J. Med.* 395:369-77, 420-25
 39. Glueck, C. J., Connor, W. E., 1978. Diet—coronary heart disease relationships reconnoitered. *J. Clin. Nutr.* 31:727-37
 40. Brunzell, J. D., Chait, A., Bierman, E. L. 1978. Pathophysiology of lipoprotein transport. *Metabolism* 27:1109-27
 41. Shaper, A. G., Lee, K. T., Scott, R. F., Goodale, F., Thomas, W. A. 1962. Chemico-anatomic studies in the geographic pathology of arteriosclerosis: comparison of adipose tissue, fatty acids and plasma lipids in diabetics from East Africa and the U.S.A. with different frequencies of myocardial infarction. *Am. J. Cardiol.* 10:390-99
 42. Glueck, C. J., Mattson, F., Bierman, E. L. 1978. Diet and coronary heart disease: another view. *N. Engl. J. Med.* 298:1471-74
 43. Schlierf, G., Stossberg, V. 1970. Diurnal patterns of plasma triglyceride, free-fatty acid, blood sugar and insulin levels on high-fat and high-carbohydrate diets in normals and in patients with primary endogenous hypertriglyceridemia. In *Atherosclerosis Proc. 2nd Int. Symp.*, ed. R. J. Jones, pp. 459-63. New York: Springer-Verlag
 44. Stone, D. B., Connor, W. E. 1963. The prolonged effects of low-cholesterol, high-carbohydrate diet upon the serum

- lipids in diabetic patients. *Diabetes* 12:127-32
45. Weinsier, R. L., Seeman, A., Herrera, G., Assal, J. P., Soeldner, J. S. 1974. High- and low-carbohydrate diets in diabetes mellitus. Study of effects on diabetic control, insulin secretion and blood lipids. *Ann. Intern. Med.* 80:332-41
 46. Ginsberg, H., Olefsky, J., Kimmerling, G., Crapo, P., Reaven, G. M. 1976. Induction of hypertriglyceridemia by a low-fat diet. *J. Clin. Endocrinol. Metab.* 42:729-35
 47. Anderson, J. W., Ward, K. 1978. Long-term effects of high-carbohydrate, high-fiber diets on glucose and lipid metabolism: a preliminary report on patients with diabetes. *Diabetes Care* 1:77-82
 48. Nichols, A. B., Ravenscroft, C., Lamphiear, D. E., Ostrander, L. D. 1976. Independence of serum lipid levels and dietary habits. The Tecumseh Study. *J. Am. Med. Assoc.* 236:1948-53
 49. Slater, G., Mead, J., Dhopeswarkar, G., Robinson, S., Alfin-Slater, R. B. 1976. Plasma cholesterol and triglyceride in men with added eggs in the diet. *Nutr. Rep. Int.* 14:249-60
 50. Connor, W. E., Lin, D. S. 1974. The intestinal absorption of dietary cholesterol by hypercholesterolemic (Type II) and normocholesterolemic humans. *J. Clin. Invest.* 53:1062-70
 51. Wynder, E. L., Reddy, B. S. 1975. Dietary fat in colon cancer. *J. Natl. Cancer Inst.* 54:7-10
 52. Siperstein, M. D., Foster, D. W., Knowles, H. C. Jr., Levine, R., Madison, L. L., Roth, J. 1977. Control of blood glucose in diabetic vascular disease. *N. Engl. J. Med.* 296:1060-65
 53. National Academy of Sciences/National Research Council: Recommended Dietary Allowances. 1974. Washington, DC. 8th ed.
 54. Huttenen, J. K., Karro, K., Makinen, K., Scheinin, A. 1975. Effects of sucrose, fructose and xylitol in glucose, lipid and urate metabolism. *Acta Odont. Scand. Suppl.* 70, pp. 239-45
 55. Miranda, P. M., Horowitz, D. L. 1978. High-fiber diets in the treatment of diabetes mellitus. *Ann. Intern. Med.* 88:482
 56. Talbot, J. M. 1978. The need for special foods and sugar substitutes by individuals with diabetes. *Life Sci. Rev. Off. FASEB.* 47 pp.

THE PROLONGED QT INTERVAL

◆7313

*J. A. Abildskov, M. D.*¹

Nora Eccles Harrison Cardiovascular Research and Training Institute
and Division of Cardiology, Department of Internal Medicine,
University of Utah College of Medicine, Salt Lake City, Utah 84132

The QT interval of the electrocardiogram (ECG) reflects the duration of electrical systole. The measurement includes the duration of the QRS complex as well as that of the ST-T deflection; however, the clinical significance of the measurement is as an index of the time required for ventricular repolarization. The duration of ventricular excitation reflected by the QRS complex is properly included in the index, since repolarization begins immediately after excitation in individual cardiac cells. The QT interval is therefore an index of the time required for repolarization from its onset in one portion of the ventricles to its completion in all portions of the ventricles. Since the duration of ventricular recovery is affected by a variety of disease states and pharmacologic agents and also has important relations to cardiac arrhythmias, study of the length of the QT interval has substantial clinical utility.

This article describes the physiologic basis of the QT interval with special attention to effects of the autonomic nervous system. The association of prolongation of the QT interval and several significant disease and drug-induced states is briefly considered. Finally, the syndromes of idiopathic QT interval prolongation and their pathophysiology and the more general prognostic significance of QT interval prolongation are considered.

¹Supported in part by Program Project Grant HL 13480 from the National Institutes of Health and the Richard A. and Nora Eccles Harrison Fund for Electrocardiographic Research.

PHYSIOLOGIC BASIS

The QT interval of the body surface electrocardiogram is related to the duration of ventricular transmembrane action potentials at the cellular level. The interval is also related to the sequence of ventricular excitation that determines the sequence of onset (and hence completion) of ventricular repolarization. In addition, the QT interval is related to the degree to which the ventricular excitation sequence and the duration of cellular transmembrane action potentials are reflected in body surface ECG leads. If these items were perfectly represented in the ECG, the QT interval would express the time between earliest ventricular excitation (and therefore earliest recovery) and completion of the latest transmembrane action potential downstroke. In a particular ECG lead, however, the onset of the QRS complex is unlikely to reflect the earliest movement of ventricular activation. Similarly, completion of the T wave in that lead may not correspond to the actual completion of ventricular recovery. It can therefore be expected that multiple simultaneous leads will provide the best electrocardiographic index of the time required for ventricular recovery.

The best index of total ventricular recovery time would be the interval between onset of the QRS complex in the lead in which this occurs earliest and completion of the T wave in the lead in which that event is latest. Even multiple simultaneous leads do not, however, insure a direct relation between the time required for ventricular recovery and the QT interval, measured as described, in the body surface ECG. Many of the possible electrocardiographic expressions of both ventricular excitation and ventricular recovery are cancelled by the multidirectional nature of these processes. It has been estimated that 70–90% of the possible ECG effects of normal ventricular excitation boundaries are cancelled and do not influence QRS waveform in leads distant from the heart (1, 2). The degree of cancellation of electrocardiographic effects of ventricular repolarization boundaries has been estimated as 92–97% for distant electrocardiographic leads (3).

Leads that include electrodes near the heart are less subject to cancellation of electrocardiographic effects by opposing directions of activation or recovery boundaries. It is unlikely, however, that any body surface leads are free of the cancelling effects of opposing boundaries of potential difference. Since some of the possible electrocardiographic expressions of both excitation and recovery boundaries are cancelled, the QT interval may not correctly reflect the time required for ventricular recovery even when determined from multiple simultaneous leads. As will be described in more detail with reference to the autonomic nervous system, it is particularly likely that late portions of ventricular recovery following normal excitation have cancelling electrocardiographic effects and are not expressed in the T wave or the duration of the QT interval (4).

In addition to cancellation of ECG effects of ventricular excitation and recovery, differences in the nature of these processes affect the significance of the QT interval. In comparison to ventricular excitation, repolarization in the ventricles is a relatively slow process. Multiple recovery boundaries between cardiac areas in different stages of repolarization coexist during most phases of recovery (5-8). In normal ventricles, the duration of intrinsic recovery properties is inversely related to the normal excitation sequence. Areas of the ventricles normally activated early have long intrinsic recovery properties, while those areas normally excited late have shorter recovery properties. With this distribution of recovery properties, portions of the ventricle with long recovery properties are capable of contributing to both early and late portions of the T wave (9). During early moments of recovery, transmembrane action potentials in cardiac areas with longer recovery properties will be in the plateau phase of transmembrane action potentials, while adjacent areas with shorter recovery properties will have action potentials in the downstroke phase. A boundary of potential difference will then exist between ventricular areas with long and short recovery properties, and this boundary will contribute to early portions of the T wave. During late moments of recovery, the downstrokes of transmembrane action potentials will be nearing completion in parts of the ventricle with long recovery properties while transmembrane action potentials in the remainder of the ventricles will have returned to resting levels. A boundary of potential difference will still exist between the area of the ventricle with long recovery properties and other portions of the ventricles and will therefore contribute to late as well as early portions of the T wave. A range of prolongation or shortening of recovery properties in areas other than those contributing to terminal portions of the T wave may thus occur without affecting the duration of the QT interval in the electrocardiogram.

For the reasons described, namely cancellation of substantial portions of the ECG effects of excitation and recovery and lack of sensitivity to local recovery properties, it is important to recognize that the QT interval is an electrocardiographic measurement and not the direct expression of a physiologic property of the heart. The interval frequently provides a useful index of the time required for ventricular recovery, but there are important exceptions, considered below in relation to autonomic nervous system effects and the idiopathic prolonged QT interval syndromes.

The QT interval is markedly influenced by heart rate, and evaluation of the significance of the interval requires consideration of this factor. Several formulae describing the relation of heart rate and the normal QT interval have been proposed on the basis of empiric observations. One of the most widely used is that of Bazett (10) in which the normal QT interval for particular heart rates is calculated as the square root of the cycle length multiplied by a constant of 0.37 seconds for men and children and 0.40

seconds for women. Tables showing normal QT interval durations for various heart rates based on Bazett's or other formulae are widely available and useful for recognizing abnormal intervals.

It should be noted that the QT interval cannot always be accurately measured. During tachycardia or in the presence of prolonged QT interval, the end of the T wave may be superimposed on the following P wave. Furthermore, the slope of the terminal portion of the T wave may be so gradual as to make determination of its end ambiguous. Another difficulty in measurement may be differentiation of T and U waves. Prominent U waves are frequent in states in which prolongation of QT intervals occur and may have a similar clinical significance. Nevertheless, inclusion of the U wave in the measurement of the QT interval is erroneous if that interval is evaluated by standards that do not include U-wave duration. Despite the difficulties of measurement, the QT interval is clearly abnormal enough in a sufficient number of patients, and when abnormal has sufficient significance, that the measurement has definite clinical utility.

AUTONOMIC NERVOUS SYSTEM INFLUENCES

Effects of the autonomic nervous system on the QT interval further elucidate the physiologic basis of the normal interval and the mechanisms of some abnormalities of the interval, and have important implications for some disease states. Vagal effects on the QT interval are largely, if not exclusively, indirect and related to parasympathetic effects on heart rate. The sympathetic nervous system has marked effects on the time required for ventricular recovery and complex but systematic effects on the ST-T deflection and QT interval of the electrocardiogram.

In experiments on dogs, Yanowitz et al (11) demonstrated prolonged QT interval and increased, positive T-wave amplitude in a longitudinal body surface lead by either right-sided stellate ganglionectomy or left stellate ganglion stimulation. Left stellate ganglionectomy or right stellate ganglion stimulation resulted in negativity of the T waves without alteration of the QT interval. The changes of T waveform were explicable on the basis of localized cardiac distribution of fibers from right and left stellate ganglia. This distribution was determined by changes of refractory periods after stellate ganglionectomy. Following right stellate ganglionectomy, ventricular refractory period prolongation occurred and was most marked in the anterior ventricular walls, while left stellate ganglionectomy produced the greatest prolongation of refractory periods in the posterior ventricular wall. Since that data indicated that left stellate ganglion fibers innervated mainly the posterior ventricular wall, left stellate stimulation resulted in a cardiac state with relatively shorter recovery times in the posterior as compared to

the anterior wall. A similar state of relative recovery times resulted when right stellate ganglionectomy prolonged recovery times on the anterior wall. Both states were characterized by earlier recovery of the posteroinferior ventricular wall, which explained the observed increase in T-wave positivity in a vertical ECG lead with the usual polarity. In a similar fashion, right stellate ganglion stimulation and left stellate ganglion ablation produced cardiac states with similar temporal relations of relative recovery times in anterior and posteroinferior ventricular walls. In those conditions, ventricular recovery times were earlier on the anterior than on the posteroinferior wall, and negative T waves in the vertical ECG lead occurred.

Although the changes of T-wave amplitude and polarity with stellate stimulation and ablation were explicable on the basis of the cardiac distribution of nerve fibers, the prolongation of the QT interval by left-sided sympathetic stimulation was unexpected. Recovery properties measured in terms of refractory periods are known to shorten with stellate stimulation. It has been postulated that the phenomenon described as cancellation (see the previous section) is responsible for paradoxical QT interval prolongation with left sympathetic stimulation. More specifically, it has been proposed that reduction of recovery times in at least part of the ventricles innervated by the left stellate ganglion unmasks previously cancelled ECG evidence of ventricular repolarization (11).

PROLONGATION DUE TO DISEASE OR DRUGS

Prolongation of the QT interval is one of many nonspecific ECG abnormalities. QT may be prolonged in virtually all varieties of organic heart disease, including ischemic disease, and in a number of other states. Like other nonspecific ECG abnormalities, the significance of QT interval prolongation requires evaluation in individual patients. When other causes of the abnormality have been excluded, the finding may suggest the presence of organic heart disease, but it is not useful in its differential diagnosis.

Prolongation of the QT interval also occurs in electrolyte disorders including hypocalcemia, hypokalemia, and hypomagnesemia. In hypocalcemia, prolongation of the QT interval is often the result of prolonged ST segments, and the T-wave duration itself may be normal. In hypokalemia and hypomagnesemia, QT interval prolongation is likely to include prolongation of the T wave. Quinidine, procainamide, and the phenothiazine derivatives frequently result in prolongation of the QT interval. Prolongation of the QT interval may also be of neurogenic origin and has been noted after cerebrovascular accidents, head trauma, central nervous system tumors and infection, and during and following neurosurgical procedures (12-22). In some cases, the electrocardiographic findings of neurogenic

origin may be characteristic and include very large T waves of either normal or abnormal polarity and prominent U waves in addition to prolongation of the QT interval.

FAMILIAL LONG QT INTERVAL SYNDROMES

The association of QT interval prolongation, congenital deafness, syncope, and sudden death was first reported by Jervell & Lange-Nielson in 1957 (23). Later, Romano et al (24) and Ward (25, 26) reported an otherwise similar syndrome but one that occurred in patients with normal hearing. Several hundred cases of the syndromes have now been reported.

The major clinical manifestation of the syndromes is recurrent syncope. The frequency of syncope is highly variable; it occurs several times daily in some patients and at intervals of years in others. Episodes of syncope usually begin early in life, although the onset may be as late as the second and third decades. There is a tendency for the frequency of syncopal attacks to decrease with increasing age. Factors that precipitate syncopal attacks include exercise, anxiety, fear, and sudden noises. Although ventricular asystole has been reported, the usual mechanism of syncopal episodes appears to be ventricular fibrillation. Sudden death is frequent in affected patients. The prevalence of the syndromes is uncertain, and it is suspected by some that the diseases are often unrecognized. Cases of the syndrome without deafness outnumber those with this manifestation by approximately 7 to 3. It should be noted that not all patients with syncope, prolonged QT intervals, and sudden death have evidence that the disorder is familial. For example, Froggatt & James (27) reported a case in which no evidence of a related disorder was found in 342 of the patients' relatives.

The electrocardiogram of patients with these syndromes shows prolongation of the QT interval with broad T waves, which may be notched, biphasic, or peaked and may be of normal or abnormal polarity. Beat-to-beat differences of T waveform may be present, and Schwartz et al (28) described episodes of T-wave alterations following stimuli similar to those that precipitate syncopal attacks. The QT interval may differ from one time to another in the same individual. Exercise usually further prolongs the QT interval.

The pathophysiology of the idiopathic prolonged QT interval syndromes has not been fully established. Abnormalities of the vasculature of the sinus node were reported by James, and disorders of myocardial metabolism were considered by some observers (29, 30). There is considerable evidence that abnormalities of the autonomic nervous system, and specifically the sympathetic portion of that system, are involved. For example, episodes of syncope in the syndromes are often precipitated by events associated with increased sympathetic nervous activity including exercise, anxiety, or fear.

Further, the most effective therapy of the syndromes involves measures that reduce the level of sympathetic nervous system influence on the heart, such as beta-blocking drugs or stellate ganglion blockade or ablation. In addition, the ECG features of QT interval prolongation and alteration of T waveform can be produced by asymmetrical alteration of sympathetic tone in experimental animals (11). In one reported case of the idiopathic prolonged QT interval syndrome, left stellate ganglion blockade reduced the QT interval to normal, and surgical removal of the left stellate ganglion abolished the previously frequent episodes of syncope (31). Subsequent observations of other patients have shown inconsistent effects of left stellate ganglion block and stellectomy on the QT interval, but surgical removal of the left stellate and first thoracic ganglion has usually been effective in eliminating or markedly reducing the frequency of syncopal attacks despite failure to normalize the QT interval (28).

Although several findings suggest that asymmetric sympathetic influence on the heart is involved in the pathogenesis of the idiopathic prolonged QT interval syndromes, the exact nature of this influence is not certain. Schwartz et al (32) suggested that abnormally low right-sided sympathetic activity may be present in some patients, and they reported experimental evidence that right stellate ganglion ablation produces a significant decrease in ventricular fibrillation threshold. These findings are compatible with the prolonged QT interval and vulnerability to ventricular arrhythmias characteristic of the clinical syndromes.

Asymmetric adrenergic cardiac influences could originate in the central nervous system, sympathetic nerves (including their distribution to the ventricle), or within the myocardium. James et al (33) reported evidence of local neuritis and neural degeneration within the sinus and AV nodes, His bundle, and ventricular myocardium in the hearts of eight patients who died suddenly and who had had syncopal attacks and long QT intervals. The etiology of the intracardiac neural disease was uncertain, but chronic viral disease or a noninfectious degenerative process were considered plausible possibilities. These authors point out that if such damaged nerves have a pattern such that neural cardiac tone can be balanced by unilateral stellate blockade or ablation, such procedures might be expected to be beneficial. They also caution, however, that random distribution of neural damage might result in states that are not improved or may even be worsened by unilateral reduction of sympathetic tone.

PROGNOSTIC SIGNIFICANCE

Prolonged QT intervals in the syndromes described in the previous section certainly have prognostic significance, since the incidence of syncope, seri-

ous arrhythmias, and sudden death in these states is high. Other observations suggest that prolongation of the QT interval has a wider range of prognostic utility. States with prolongation of the QT interval due to drugs or electrolyte disorders are well known to be associated with the risk of sudden death. Transient prolongation of the QT interval has been observed in experimental animals following the rapid intravenous injection of epinephrine or norepinephrine (4). This prolongation corresponds in time to cardiac conditions of increased disparity of repolarization properties and enhanced vulnerability to ventricular fibrillation. In patients with myocardial infarction, prolongation of the QT is associated with a substantially increased risk of sudden death (34). Hynes and co-workers (35) analyzed the QT interval of patients with ischemic heart disease who had been successfully treated for ventricular fibrillation outside the hospital. Thirty-seven percent of these patients showed prolongation of the QT interval while only 18% of patients with myocardial infarction uncomplicated by ventricular fibrillation had such prolongation. The relation of QT interval prolongation and occurrence of sudden death in such disease states as the idiopathic long QT interval syndromes, drug- and electrolyte-induced conditions, and ischemic heart disease suggests the electrocardiographic finding may be a nonspecific but useful predictor of the likelihood of sudden death.

Literature Cited

1. Abildskov, J. A., Klein, R. M. 1962. Cancellation of electrocardiographic effects during ventricular excitation. *Circ. Res.* 11:247-51
2. Schaefer, H. 1952. Über die absolute Grösse elektrokardiographischer Potentiale. *Pflügers Arch. Gesamte Physiol. Menschen Tiere* 255:251-72
3. Burgess, M. J., Millar, K., Abildskov, J. A. 1969. Cancellation of electrocardiographic effects during ventricular recovery. *J. Electrocardiol.* 2:101-7
4. Abildskov, J. A. 1976. Adrenergic effects on the QT interval of the electrocardiogram. *Am. Heart J.* 92:210-16
5. Burgess, M. J., Green, L. S., Millar, C. K., Wyatt, R. F., Abildskov, J. A. 1972. The sequence of normal ventricular recovery. *Am. Heart J.* 84:660-69
6. Abildskov, J. A., Rush, S., Burgess, M. J., Millar, K. 1972. Graphic studies concerning the body surface distribution of T potentials. *Am. Heart J.* 83:283-84
7. Burgess, M. J., Millar, K., Abildskov, J. A. 1970. The geometry of ventricular repolarization boundaries. *Am. Heart J.* 79:524-30
8. Abildskov, J. A. 1975. The sequence of normal recovery of excitability in the dog heart. *Circulation* 52:422-46
9. Harumi, K., Burgess, M. J., Abildskov, J. A. 1966. A theoretic model of the T wave. *Circulation* 34:657-68
10. Bazett, H. C. 1920. An analysis of the time-relation of electrocardiograms. *Heart* 7:353-70
11. Yanowitz, F., Preston, J. B., Abildskov, J. A. 1966. Functional distribution of right and left stellate innervation to the ventricles. Production of neurogenic electrocardiographic changes by unilateral alteration of sympathetic tone. *Circ. Res.* 18:416-28
12. Byer, E., Ashman, R., Toth, L. A. 1947. Electrocardiograms with large upright T waves and long Q-T intervals. *Am. Heart J.* 33:696-806
13. Burch, G. E., Meyers, R., Abildskov, J. A. 1954. A new electrocardiographic pattern observed in cerebrovascular accidents. *Circulation* 9:719-23
14. Hayashi, S., Watanabe, J., Miyagawa, S., Tamakuma, S., Nagaki, H. 1961. Studies of electrocardiographic patterns in cases with neurosurgical lesions. *Jpn. Heart J.* 2:92-111

15. Hersch, C. 1961. Electrocardiographic changes in head injuries. *Circulation* 23:853-60
16. Greenhoot, J. H., Reichenbach, D. D. 1969. Cardiac injury and subarachnoid hemorrhage. A clinical, pathological and physiological correlation. *J. Neurosurg.* 30:521-31
17. Finkelstein, D., Nigaglioni, A. 1961. Electrocardiographic alterations after neurosurgical procedures. *Am. Heart J.* 62:772-84
18. Shuster, S. 1960. The electrocardiogram in subarachnoid haemorrhage. *Br. Heart J.* 22:316-20
19. Fentz, V., Gormsen, J. 1962. Electrocardiographic patterns in patients with cerebrovascular accidents. *Circulation* 25:22-28
20. Hugenholtz, P. G. 1967. Electrocardiographic changes typical for central nervous system disease after right radical neck dissection. *Am. Heart J.* 74: 438-41
21. Hugenholtz, P. G. 1962. Electrocardiographic abnormalities in cerebral disorders. Report of six cases and a review of the literature. *Am. Heart J.* 62:451-61
22. Grossman, M. A. 1976. Cardiac arrhythmias in acute central nervous system disease. Successful management with stellate ganglion block. *Arch. Intern. Med.* 136:203-7
23. Jervell, A., Lange-Nielsen, F. 1957. Congenital deaf mutism, functional heart disease with prolongation of the Q-T interval and sudden death. *Am. Heart J.* 54:59-68
24. Romano, C., Gemme, G., Pongiglione, R. 1963. Aritmie cardiache rare dell'eta' pediatrica. II. Accessi sincopali per fibrillazione ventricolare parossistica. *Clin. Pediatr.* 45:656-83
25. Ward, O. 1963. *Rep. Counc. R. Acad. Med. Ireland.*
26. Ward, O. C. 1964. A new familial cardiac syndrome in children. *J. Ir. Med. Assoc.* 54:103-6
27. Froggatt, P., James, T. N. 1973. Sudden unexpected death in infants. Evidence on a lethal cardiac arrhythmia. *Ulster Med. J.* 42:136-52
28. Schwartz, P. J., Periti, M., Malliani, A. 1975. The long Q-T syndrome. *Am. Heart J.* 89:378-90
29. James, T. N. 1969. QT prolongation and sudden death. *Mod. Concepts Cardiovasc. Dis.* 38:35-37
30. Garza, L. A., Vick, R. L., Nora, J. J., McNamara, D. G. 1970. Heritable Q-T prolongation without deafness. *Circulation* 41:39-48
31. Moss, A. J., McDonald, J. 1971. Unilateral cervicothoracic ganglionectomy for the treatment of long Q-T interval syndrome. *N. Engl. J. Med.* 285:903-4
32. Schwartz, P. J., Snebold, N. G., Brown, A. M. 1976. Effects of unilateral cardiac sympathetic denervation on the ventricular fibrillation threshold. *Am. J. Cardiol.* 37:1034-40
33. James, T. N., Froggatt, P., Atkinson, W. J., Luria, P. R., McNamara, D. G., Miller, W. W., Schloss, G. T., Carroll, J. F., North, R. L. 1978. Observations on the pathophysiology of the long QT syndromes with special reference to the neuropathology of the heart. *Circulation* 57:1221-31
34. Schwartz, P. J., Wolf, S. 1978. QT interval prolongation as a predictor of sudden death in patients with myocardial infarction. *Circulation* 57:1074-77
35. Haynes, R. E., Hallstrom, A. P., Cobb, L. A. 1977. Repolarization abnormalities in sudden cardiac death syndrome. *Clin. Res.* 25:143a (Abstr.)

COMPUTED TOMOGRAPHY OF THE HEPATOBILIARY SYSTEM

◆7314

Melvyn Korobkin, M.D. and Henry I. Goldberg, M.D.

Department of Radiology, University of California, School of Medicine,
San Francisco, CA 94143

INTRODUCTION

Computed tomography (CT) has truly revolutionized the field of diagnostic radiology. In the period of six years since its introduction into clinical practice, CT scanning has become the indispensable focal point of the diagnostic evaluation of a suspected structural abnormality of the brain. During the past three years, development of CT scanners capable of completing a scan while the patient suspends respiration has led to initial experience with CT evaluation of extracranial structures, including the hepatobiliary system. This review describes and illustrates the CT images of normal and abnormal liver and biliary tract, and it addresses the more difficult question of how CT should relate to the other available diagnostic modalities.

BASIC PRINCIPLES

Computed tomography (CT) is a new imaging modality utilizing a highly collimated X-ray beam and a high speed minicomputer. Using the data obtained by multiple X-ray detectors, the computer reconstructs a transverse cross-sectional image or tomogram as an array or grid of small picture elements and displays it as a picture on a television screen. The numerical value assigned to each grid element is a close approximation to the X-ray attenuation coefficient of the tissue element it represents. In most cases, the attenuation value can be considered to approximate the physical density of the tissue. Tissues with the highest density, such as those containing bone or iodine, appear white on the television screen; tissues of low density, such

as air or fat, have a black or dark grey appearance. Unlike conventional radiography, CT is capable of displaying and resolving tissues having density differences of the order of only one percent.

NORMAL ANATOMY

The liver is the predominant structure on CT cross sections of the upper abdomen (Figure 1). The normal liver parenchyma has a homogeneous density, equal to or slightly higher than that of the other organs of the upper abdomen, and has a smooth contour penetrated by several normal fissures. The size and shape of the normal liver varies from patient to patient; the left lobe is more variable than the right. Intrahepatic portions of the inferior vena cava or portal vein are occasionally identified as round or curvilinear structures of lower density than the surrounding liver parenchyma, but normal intrahepatic bile ducts are not visualized with currently available CT scanners. The gallbladder is normally seen as an oval, low density structure along the inferomedial border of the right lobe of the liver. Although gallstones are occasionally identified, CT has not been found to be useful in the evaluation of gallbladder disease. A normal common hepatic duct (CHD) may occasionally be visualized in the porta hepatis, especially



Figure 1 CT scan of the normal upper abdomen. Note the homogeneous texture of the liver (L) and its density similar to that of the spleen (S).

following the intravenous injection of standard urographic contrast material: The density of the CHD remains near that of water, while the portal veins and hepatic arteries increase in density because of the iodinated contrast material circulating within them.

MEDICAL DISEASES OF THE LIVER

Experience with CT of medical diseases of the liver is small, but most reports (1-3) indicate that hepatocellular disease cannot be identified unless marked alteration in the size or shape of the liver is present, or the liver is infiltrated by large amounts of fat (Figure 2) or iron. Most radiologists using CT body scanning have observed patients with alcoholic liver disease who demonstrate a liver with a uniform decrease in density of the liver parenchyma, but the sensitivity of this finding and its precise relationship to the amount of fatty infiltration is unknown. Preliminary animal studies in our laboratory indicate that the addition of only 5% of neutral fat to the liver mass can produce a remarkable reduction in CT attenuation value of the liver.

It is likely that patients with hemochromatosis can be identified by CT. Twelve of 14 patients clinically thought to be iron overloaded demonstrated



Figure 2 Fatty infiltration of the liver. Presence of fat has resulted in a lower density than the normal spleen.

livers of higher density than 12 hospital controls (4). The precise relationship between hepatic iron and the density as determined by CT is unknown. A fruitful subject for future research is the potential utility of CT as a method of following the course of treatment of both hemochromatosis and fatty infiltration of the liver.

MASS LESIONS OF THE LIVER

A mass lesion within the liver can be detected by CT if its X-ray attenuation coefficient (density) is sufficiently different from the density of surrounding liver parenchyma. Almost all liver masses detected by CT have a density less than normal liver; the exceptions are those occasional lesions containing calcium or fresh hematoma (2). Cysts of the liver are usually easy to identify because their density is near that of water, easily distinguishable from the higher density of normal liver (Figure 3). Neoplasms of the liver, whether primary or metastatic, usually have attenuation values higher than cysts but somewhat lower than normal surrounding liver. Abscesses and hematomas

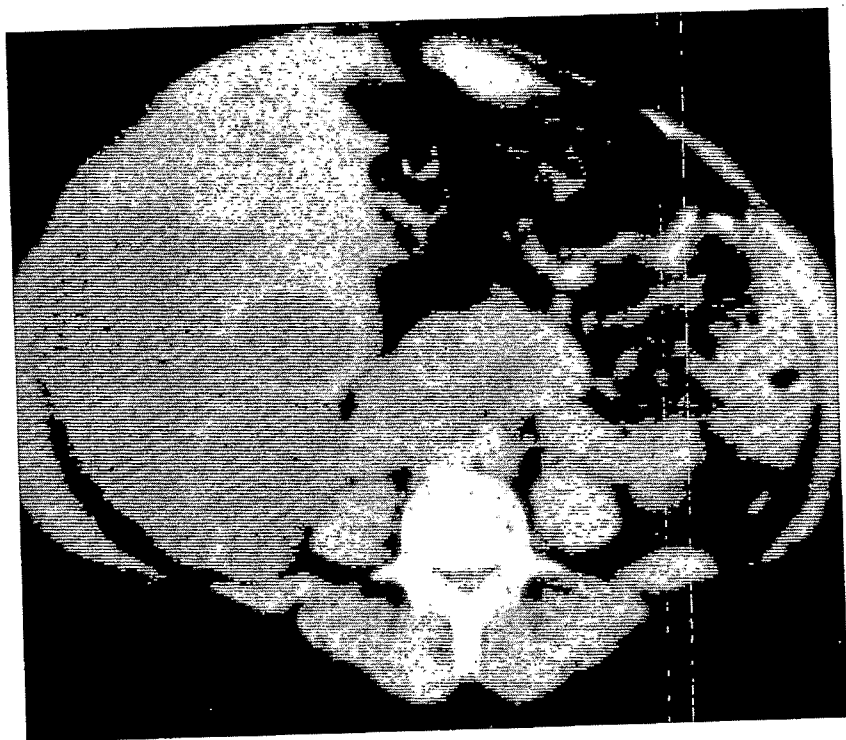


Figure 3 Cystic disease of the liver. Several low density, circular masses are present in the posterior aspect of the right lobe of the liver. Density of these lesions approximated that of water, correctly suggesting the diagnosis of cysts rather than metastases.

of the liver have densities intermediate in value between cysts and neoplasms. Detection of subtle density differences between tumors and surrounding liver often requires special attention to manipulation of the electronic window settings on the scanner. The attenuation value of one large hepatoma was so similar to surrounding normal parenchyma that it was not detected by CT (2). Injection of intravenous urographic contrast material usually facilitates detection of a mass lesion by perfusing the surrounding normal liver parenchyma to a greater degree than the less vascular lesion. On occasion, however, a highly vascular lesion will become isodense with the surrounding hepatic parenchyma after injection of contrast agent and will no longer be apparent. Most investigators recommend that CT examinations of the liver be performed both before and after injection of contrast agent.

JAUNDICE

Differentiation of obstructive from nonobstructive jaundice can be made using CT. A branching pattern of curvilinear structures in the liver usually indicates the presence of dilated intrahepatic bile ducts (Figure 4). Because the normal portal venous tree occasionally produces a similar appearance within the liver, intravenous injection of contrast material is sometimes



Figure 4 Obstructive jaundice. The curvilinear and oval, low density structures (arrows), in the liver, represent dilated intrahepatic bile ducts.

necessary to distinguish bile ducts from blood vessels (5). Dilated biliary structures will remain visible as low density regions, but vascular structures will either become isodense (and disappear) or attain a higher density than surrounding liver. Injection of urographic contrast media is also useful in searching for dilated extrahepatic biliary ducts. The increased density of vascular structures in the porta hepatis and of the head of the pancreas that results from contrast injection makes it easier to identify a low density, dilated common hepatic duct or common bile duct (Figure 5). Evaluation of the extrahepatic biliary tree is particularly important in jaundiced patients if there is no evidence of low density, curvilinear structures in the liver; we recently examined six patients with obstructive jaundice whose CT scans showed a dilated extrahepatic biliary tree in the absence of intrahepatic ductal dilatation (6).

Several studies have indicated a high degree of accuracy in predicting the status of the biliary tree in jaundiced patients using CT. Havrilla et al (7) reported 85% accuracy in identifying obstructive jaundice (22 of 25 patients) and correctly diagnosed no obstruction in the remaining 19 patients in their series. Levitt and co-workers (3) made a correct diagnosis in 44 of 45 patients with obstructive jaundice; 41 of their 46 patients with nonobstructive jaundice were correctly identified by CT. In our own prospective



Figure 5 Obstructive jaundice. Dilatation of the distal common bile duct is visualized as a circular, low density structure (arrow) in the head of the pancreas.

study of 23 jaundiced patients (8), CT correctly indicated the presence or absence of obstruction in 20; a retrospective study of the same material increased the accuracy to 22 of 23 patients because of recognition of common bile duct dilatation in the absence of intrahepatic duct dilatation. CT is useful not only in the detection of a dilated biliary tree, but also in the identification of the cause of biliary obstruction. The most common cause of extrahepatic biliary obstruction, visualized by CT, is a carcinoma of the pancreas, but occasionally a stone in the common bile duct, or enlarged lymph nodes in the porta hepatis, will be visualized.

RELATIONSHIP TO OTHER IMAGING MODALITIES

Because CT is an expensive new technology, there is widespread interest in a careful assessment of the appropriate role for CT in the diagnostic approach to each disease process for which it is being used. Unfortunately, funding for properly controlled, prospective studies of the relative usefulness of extracranial CT has not been available. A rigorous, prospective comparison of CT, ultrasound, and radionuclide imaging of an unselected group of patients with suspected hepatobiliary disease has not been performed. In addition to obtaining funds to carry out such a study, it would be necessary to utilize state-of-the-art equipment and have a skilled and experienced team of technologists and radiologists—a difficult task given the new and rapidly changing technologies of CT and ultrasound.

One large retrospective comparison of CT and radionuclide imaging (RI) stressed the following preliminary conclusions (9):

1. The overall sensitivity of the two techniques in detecting focal mass lesions is similar, about 85–90%.
2. CT is quite helpful in evaluation of those areas known to create problems for RI, i.e. single defects in the region of the porta hepatis, gallbladder fossa, or left lobe. The presence or absence of a pathological basis for an abnormal colloid liver scan can usually be determined by CT.
3. CT can usually determine whether a mass lesion seen on RI study is due to cyst, abscess, hematoma, or tumor.
4. Except in patients with fatty liver or hemochromatosis, RI is clearly more sensitive in the detection of hepatocellular disease than is CT.
5. In the jaundiced patient, CT is clearly more accurate in the determination of whether the biliary tract is obstructed.

Comparisons of ultrasound and CT are even more scarce than RI and CT. In our own prospective comparison of ultrasound and CT in the evaluation of jaundice (8), no real difference in accuracy was apparent. Decisions regarding which examination to use rest largely on the relative

degree of experience and skill with each modality, as well as on the quality of the apparatus available. One retrospective study indicated a superiority of ultrasonography in the detection of space-occupying lesions, but a 2½-minute CT scanner was used in that evaluation (10). Availability of a good study of all three modalities, even retrospective, is necessary before a clear idea can be gained about the relative roles of CT, RI, and ultrasound in the detection and characterization of space-occupying lesions of the liver.

CONCLUSIONS

Computed tomography is a new diagnostic modality capable of producing remarkably clear images of the hepatobiliary system. Infiltration of the liver with fat or iron produces characteristic alteration of tissue density easily detected by CT. Mass lesions can be detected and can usually be characterized as either cystic or solid in nature. Surgical jaundice is easily differentiated from medical jaundice by the visualization of a dilated intra- or extrahepatic biliary tree. The cause of surgical jaundice is often detected, especially if a mass is present in the head of the pancreas. The precise role of CT in the diagnostic evaluation of known or suspected hepatobiliary disease has yet to be determined.

Literature Cited

1. Alfidi, R. J., Haaga, J. R., Havrilla, T. R., Pepe, R. G., Cook, S. A. 1976. Computed tomography of the liver. *Am. J. Roentgenol.* 127:69-74
2. Stephens, D. H., Sheedy, P. F. II, Hattery, R. R., MacCarty, R. L. 1977. Computed tomography of the liver. *Am. J. Roentgenol.* 128:579-90
3. Levitt, R. G., Sagel, S. S., Stanley, R. J., Jost, R. G. 1977. Accuracy of computed tomography of the liver and biliary tract. *Radiology* 124:123-28
4. Mills, S. R., Doppman, J. L., Nienhuis, A. W. 1977. Computed tomography in the diagnosis of disorders of excessive iron storage of the liver. *Comput. Tomogr.* 1:101-4
5. Kressel, H. Y., Korobkin, M., Goldberg, H. I., Moss, A. A. 1977. The portal venous tree simulating dilated biliary ducts on computed tomography of the liver. *Comput. Tomogr.* 1:169-75
6. Shanser, J. D., Korobkin, M., Goldberg, H. I., Rohlfing, B. M. 1978. Computed tomographic diagnosis of obstructive jaundice in the absence of intrahepatic ductal dilatation. *Am. J. Roentgenol.* 131:389-92
7. Havrilla, T. R., Haaga, J. R., Alfidi, R. J., Reich, N. E. 1977. Computed tomography and obstructive biliary disease. *Am. J. Roentgenol.* 128:765-68
8. Goldberg, H. I., Filly, R. A., Korobkin, M., Moss, A. A., Kressel, H. Y., Callen, P. W. 1979. Capability of CT body scanning and ultrasonography to demonstrate the status of the biliary ductal system in patients with jaundice. *Radiology*. In press
9. Biello, D. R., Levitt, R. G., Siegel, B. A., Sagel, S. S., Stanley, R. J. 1978. Computed tomography and radionuclide imaging of the liver: a comparative evaluation. *Radiology* 127:159-63
10. Bryan, P. J., Dinn, W. M., Grossman, Z. D., Wistow, B. W., McAfee, J. G., Kieffer, S. A. 1977. Correlation of computed tomography, gray scale ultrasonography, and radionuclide imaging of the liver in detecting space-occupying processes. *Radiology* 124:387-93

COMPUTED TOMOGRAPHY IN NEUROLOGIC DIAGNOSIS

◆7315

Anne G. Osborn, M.D.

Department of Radiology, University of Utah College of Medicine,
Salt Lake City, Utah 84132

INTRODUCTION

The instant, universal acceptance of computed tomography (CT) in the evaluation of a wide spectrum of intracranial lesions attests to its great value in neurologic diagnosis. The reliability of CT in demonstrating a broad range of intracerebral disorders has been well established (1). CT has been said to provide more diagnostic information on the general spectrum of intracranial disease than any other single neuroradiologic examination (2) and for the first time permits direct visualization of important but formerly obscure areas such as the orbit and posterior fossa.

This report reviews the basic physical principles of CT scanning, then briefly discusses the utility, economics and efficacy, limitations, and new developments in cranial CT as an adjunct to neurologic diagnosis.

BASIC PRINCIPLES OF CT SCANNING

The ability to distinguish small differences in the x-ray attenuation of various intracranial tissues—ordinarily not distinguishable with conventional neurodiagnostic procedures—in large part accounts for the unparalleled success of CT (3). A series of narrowly collimated x-ray exposures is made of an object as viewed from many different angles, and the data thus acquired is then used to reconstruct sectional images that display the object in cross section.

The first commercially available CT scanners employed a translate-rotate data-gathering format; i.e. after one complete transverse scan, the x-ray tube and its highly sensitive detector system were rotated 1° around the object to be imaged and the scan repeated, usually through an arc of 180°.

Thus 180 different data profiles were obtained of the imaged object and the entire section was then reconstructed by a computer using a convolution algorithm or "filtered back projection" method (3, 4).

More recent developments in CT technology include fan-beam systems (where the x-ray source and detectors are rotated synchronously), fixed or stationary detector systems with a fan-beam radiation source, and (still in the future) fixed-source, fixed-detector scanners. These innovations have both improved resolution and reduced the requisite scan time from several minutes to only a few seconds.

UTILITY, EFFICACY, AND ECONOMICS

Virtually all neurodiagnostic techniques, with the possible exception of EEG, have been used less often as the capabilities of CT have become known to clinicians (5). CT scans in many instances obviate the need for other, more invasive diagnostic procedures such as cerebral angiography. In other cases, CT has added important complementary information that facilitates patient management (2).

Some authors estimate that CT is approximately 90–98% accurate in imaging most intracranial lesions and is thus one of the most effective medical detection systems available (6). The utility of CT in the assessment of intracranial trauma and/or hemorrhage, degenerative diseases, hydrocephalus, mass lesions, and congenital malformations has been thoroughly documented (7–13). CT has also proved particularly helpful for follow-up studies on patients with cerebral tumors undergoing surgery and/or radiation therapy (14), ventricular shunting for hydrocephalus (15), in evaluating the treatment course of patients with cerebral abscess (16), and for delineating lesions in the orbit, sella, and posterior fossa. CT has profoundly altered the management of posterior fossa tumors in children, many of whom have been explored surgically after investigation by CT only (17). On the other hand, CT has resulted in little demonstrable improvement either in the cost of care or management of patients with cerebrovascular disease (18).

Abrams & McNeil (19–21) pointed out that the remoteness of health outcome from the point of a diagnostic test such as CT leads to interim reliance on more proximate measures of that test's efficacy, such as (a) the degree to which CT furnishes new or additional diagnostic information not otherwise available; (b) its accuracy; (c) its effect on the morbidity and mortality of diagnostic and therapeutic procedures; (d) its impact on treatment planning; and (e) changes in cost and/or savings incident to its use.

While the lack of hard, controlled data makes evaluation of many efficacy studies hazardous, the efficacy of CT in difficult management problems

common to general neurologic practice has nevertheless been studied. For example, because of the low yield of therapeutically significant lesions, it has rarely been justifiable to subject those chronic childhood seizure patients without progressive neurological defect to the morbidity of pneumoencephalography or arteriography. As a noninvasive screening procedure, CT has disclosed structural abnormalities in 30% of such cases. While the most common abnormality detected is either focal or generalized atrophy (15%), only about 2% of the abnormalities discovered by CT are potentially of therapeutic significance (22). However, the demonstration of either a static process or findings of a normal CT may be helpful in some cases to both family and physician.

The utility of CT in the assessment of childhood macrocrania has also been documented (23). Only 18% of these sometimes puzzling patients had a normal CT. The remainder had significant hydrocephalus, neoplasms, subdural hematomas or effusions, or a variety of abnormalities involving the cerebral parenchyma.

Patients presenting with the chief complaint of chronic headache or a presumptive diagnosis of temporal lobe epilepsy form a significant portion of outpatients scheduled for CT as a primary or "screening" procedure. In patients with headache and no objective neurologic findings, CT scans are usually normal. In contrast, a recent study found 34% with headache *and* associated objective neurologic findings had abnormal CT scan (55% were regarded as potentially treatable; the majority of the others had demonstrable cortical atrophy). Assessing the cost at \$240 for each examination, the calculated lower bounds for costs of case findings in these groups were \$4,363 and \$500 per patient respectively (24).

At one university hospital, 35% of all CT scans requested resulted in a positive, clinically important diagnosis for an overall case-finding cost of \$800. Case-finding costs here varied from a low (and therefore high positive yield) of \$411 for patients in coma to \$3,500 for patients with headaches as their only indication for CT scanning (24).

Since CT became available, one study reports a progressive decrease in the proportion of cerebral angiograms that are normal (from 36 to 16%) (25). Hence CT is also making those other neurodiagnostic evaluations that are still performed more efficient for patients with suspected lesions. Preliminary data also show that in many cases, CT can replace or reduce more "conventional" diagnostic studies, resulting in significant reduction in hospital stay and cost (26), particularly in certain patient groups such as those with extracerebral collections or intracerebral neoplasms. Problems that still need to be resolved include proper utilization, further demonstration of cost-effectiveness, and evaluation of future technological developments.

COMPLEMENTARY ROLE OF CT AND OTHER NEURODIAGNOSTIC PROCEDURES

While it has replaced more invasive procedures in some instances, the complementary role of CT and additional neuroradiologic studies has been well documented in others (2). The accuracy of CT varies according to the organ system examined as well as the locus and specific nature of the lesion (27).

CT may detect blood in the subarachnoid space, intracerebral or intraventricular hematomas, the thrombosed portion of a giant aneurysm, etc. However, angiography in the patient with aneurysm and subarachnoid hemorrhage is still necessary to visualize the base of the aneurysm, its relationship to adjacent vessels, and the presence of vascular spasm.

CT can demonstrate areas of cerebral ischemia or infarction, if they are of sufficient size, and can easily distinguish between hemorrhagic and ischemic infarction. Here again, angiography is necessary to delineate precise vascular details such as vessel occlusions, stenosis, ulcerating plaques, and patterns of collateral blood flow.

CT alone is usually sufficient as a diagnostic procedure in patients with acute head injury and can accurately distinguish traumatic edema from frank intercerebral hematoma. Angiography may be necessary to delineate vessel lacerations, traumatic fistulae, and isodense (resolving) balanced subdural hematomas.

While CT is more accurate than angiography or radionuclide scans in detecting both primary and metastatic neoplasms, angiography is necessary for the delineation of feeding vessels, occluded dural sinuses, etc, and is often helpful in distinguishing between the many various entities (such as infarction) that may resemble a neoplasm (Figure 1).

Large sellar and suprasellar masses are easily detected by CT. But angiography and/or pneumoencephalography may be necessary both for differential diagnosis (Figure 2) and to delineate the exact relationship of a given mass to adjacent structures such as the optic chiasm and anterior recesses of the third ventricle. Pituitary microadenomas are also delineated best by pneumoencephalography combined with thin-section hypocycloidal tomography.

While CT has proved exceptionally helpful in the detection of posterior fossa lesions, cerebellopontine-angle tumors smaller than one centimeter are best delineated by contrast cisternography. Rarely, air studies with tomography may be necessary to distinguish between lesions such as cystically dilated fourth ventricle or cystic fourth ventricular neoplasm, which can have identical appearances on CT scan.

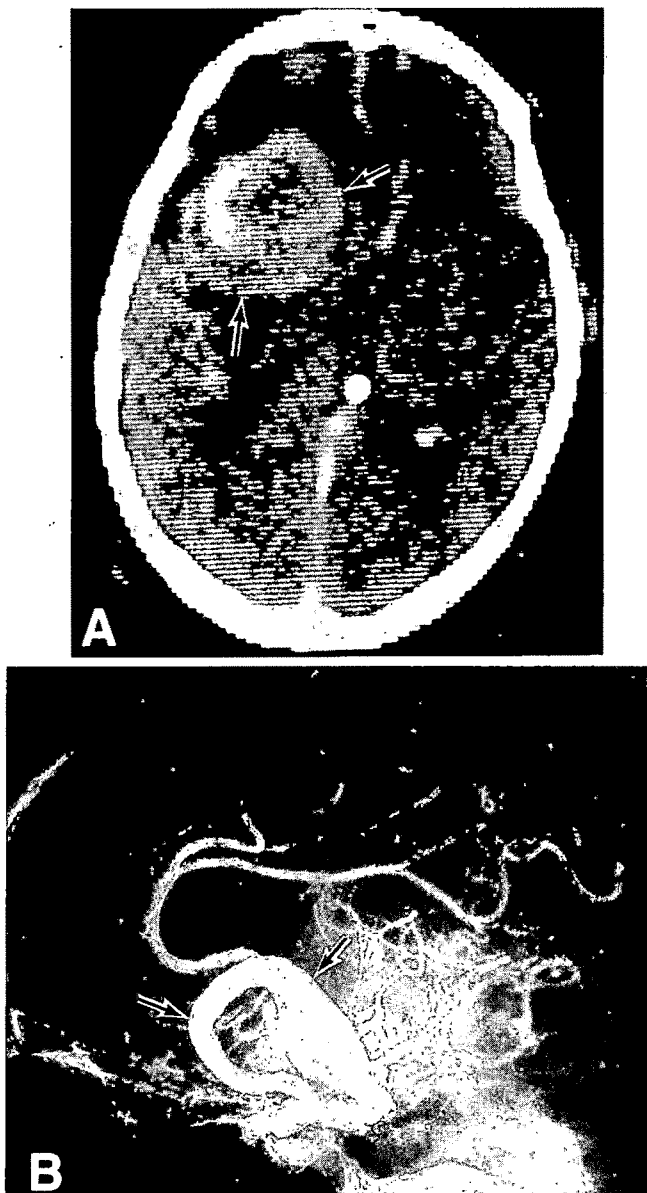


Figure 1 A. Contrast-enhanced CT scan showing a large left temporal mass lesion (*arrows*) with some peripheral edema and left to right subfalcine shift of the lateral ventricles. Does the lesion represent a tumor? B. Left common carotid angiogram, lateral view, midarterial phase. Same patient as A. A partially thrombosed giant middle cerebral artery aneurysm (*arrows*) is present, accounting for the unusual CT scan.

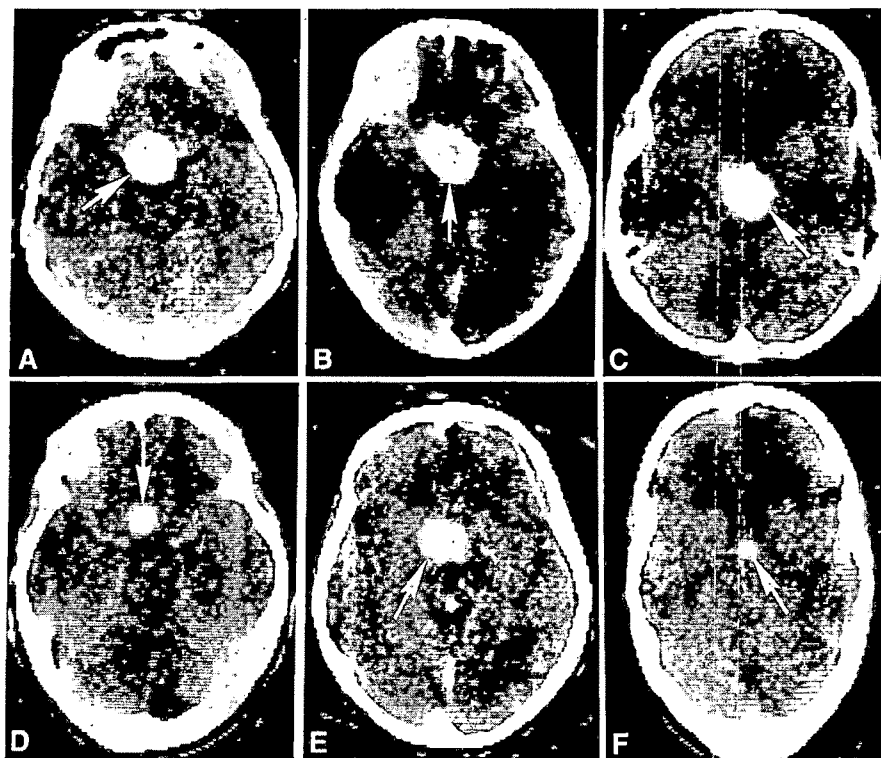


Figure 2 CT scans demonstrating a variety of contrast-enhancing lesions (arrows) in the suprasellar area. A. Giant aneurysm; B. Hypothalamic glioma; C. Craniopharyngioma, partially calcified; D. Meningioma; E. Pituitary adenoma with suprasellar extension; F. Third ventricular colloid cyst.

LIMITATIONS OF CT SCANS

While highly efficacious in many instances, various limitations are present in CT scanning (21). Some of these limitations are imposed by the system itself. Lesions below the inherent limits of resolution of the system itself either in size or attenuation difference from surrounding normal tissue may not be detected at all. Some examples include small infarcts or metastases and early infiltrating neoplasms.

Lesions may be obscured by adjacent structures, particularly at the skull base. Artifacts introduced by voluntary or biological motion can either obscure abnormalities actually present or create the appearance of non-existent lesions. Even with the use of contrast enhancement, vascular details

such as small vessel occlusions or stenosis, small to medium size aneurysms, and the vascular supply to vascular malformations or neoplasms may not be delineated.

Other limitations on CT scanning are observer imposed. Normal structures (such as a prominent jugular tubercle) may be erroneously interpreted as a lesion (acoustic neurinoma). Failure to recognize such factors as partial-volume effect (inclusion of only part of a lesion or structure in a section of finite thickness) may also lead to misdiagnosis. Failure to obtain a complete scan series, including a vertex pair, may result in failure to visualize as much as 10–15% of the total intracranial volume (28).

As the general experience with CT has increased, that virtually “anything can look like anything else” has become a caveat. Infarcts and resolving hematomas can be indistinguishable from neoplasm, neoplasms can resemble abscesses, abscesses can be identical to cavitating metastases, and metastases can mimic primary tumors (Figure 1). A variety of completely different pathologic entities in a similar location may closely resemble one another and therefore require additional diagnostic procedures (such as angiography) to assist in the radiographic differential diagnosis (Figure 2).

CONCLUSION: NEW DIRECTIONS IN CRANIAL CT SCANNING

The impending development of ultra rapid scanning systems will make both dynamic studies and gated scans of rapidly moving objects technically feasible. Currently available rapid scan units render heavy sedation or anesthesia unnecessary in virtually all uncooperative or pediatric age patients.

While computer reconstructions of frontal and lateral views of the brain from axial scan data have been performed experimentally since the advent of CT scanning, regenerated coronal and sagittal slices are now a standard feature of some commercially available units (29). With the advent of large aperture air-gap body scanners, true or nonregenerated scans with significantly improved resolution (Figure 3) can be obtained in these planes (30).

Techniques for stereotaxic CT-guided surgery are being developed (31). Relatively simple computer graphics systems for digitization of two-dimensional information and its display in three dimensions have been used experimentally (R. Brown and W. R. Glenn, personal communication) and should be available for general use in the near future.

Body CT is being utilized with increasing frequency in the evaluation of traumatic spinal injuries (32), the narrow canal syndrome, and in a variety of congenital lesions such as craniovertebral anomalies and spinal dysraph-

ism (33). The availability of metrizamide will also make CT-assisted myelography and cisternography a widely available procedure (34).

Finally, positron or emission CT scanning (ECAT) systems capable of providing high contrast, high resolution quantitative images of metabolic or physiologic parameters in the brain are being developed (35). This new technique promises to be particularly helpful in those neurologic disorders such as acute demyelinating diseases or small infarcts where routine transmission CT is often negative.

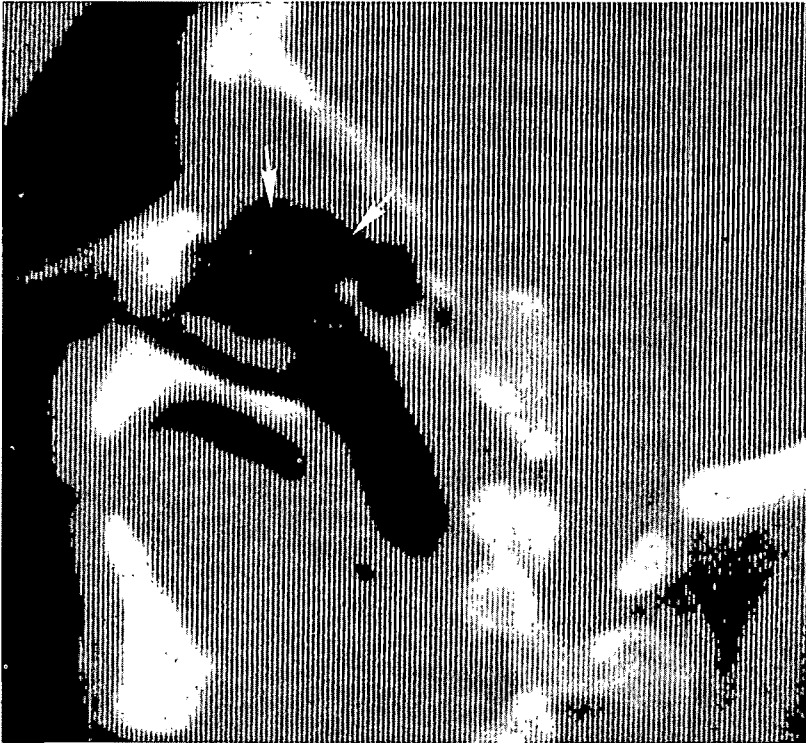


Figure 3 Direct sagittal CT scan of the face and paranasal sinuses. The scanner window width was set to optimize visualization of structures such as the ethmoid sinuses (*arrows*).

Literature Cited

1. Evens, R. G. 1976. New frontier for radiology: Computed tomography. *Am. J. Roentgenol.* 126:1117-29
2. Davis, K. R., Poletti, C. E., Roberson, G. H., Tadmor, R., Kjellberg, R. N. 1978. Complementary role of computed tomography and other neuroradiologic procedures. *Surg. Neurol.* 8:437-46
3. Ter-Pogossian, M. M. 1977. Computerized cranial tomography: Equipment and physics. *Semin. Roentgenol.* 12: 13-25
4. Brooks, R. A., DiChiro, G. 1976. Principles of computer assisted tomography (CAT) in radiographic and radioisotopic imaging. *Phys. Med. Biol.* 21:689-732
5. Baker, H. L. Jr. 1975. The impact of computed tomography on neuroradiologic practice. *Radiology* 116:637-40
6. Davis, D. O. 1977. CT in the diagnosis of supratentorial tumors. *Semin. Roentgenol.* 12:97-108
7. Baker, H. L. Jr., Campbell, J. K., Houser, O. W., Reese, D. F. 1975. Early experience with the EMI scanner for study of the brain. *Radiology* 116: 327-33
8. Harwood-Nash, D. C. 1977. Congenital craniocerebral abnormalities and computed tomography. *Semin. Roentgenol.* 12:39-51
9. Huckman, M. S., Fox, J. H., Ramsey, R. G. 1977. Computed tomography in the diagnosis of degenerative diseases of the brain. *Semin. Roentgenol.* 12:63-75
10. Davis, K. R., Taveras, J. M., Roberson, G. H., Ackerman, R. H., Dreeslath, J. N. 1977. Computed tomography in head trauma. *Semin. Roentgenol.* 12: 53-62
11. El Gammal, T., Allen, M. B. Jr., Lott, T. 1977. Computer assisted tomography and pneumoencephalography in nontumorous hydrocephalus in infants and children. *J. Comput. Assist. Tomogr.* 1:204-10
12. Zimmerman, R. A., Bilaniuk, L. A., Genneralli, T., Bruce, D., Dolinskas, C., Uzzell, B. 1978. Cranial computed tomography in diagnosis and management of acute head trauma. *Am. J. Roentgenol.* 131:27-34
13. Larson, E. B., Omenn, G. S., Magno, J. 1978. Impact of computed tomography in the care of patients with suspected hydrocephalus. *Am. J. Roentgenol.* 131:41-44
14. Kretschmar, K., Aulich, A., Schindler, E., Lange, S., Grumme, T., Muse, W. 1978. The diagnostic value of CT for radiotherapy of cerebral tumors. *Neuroradiology* 14:245-50
15. Palmieri, A., Menichelli, F., Pasquini, U., Salvatini, U. 1978. Role of computed tomography in the postoperative evaluation of infantile hydrocephalus. *Neuroradiology* 14:257-62
16. Whelan, M. A., Hilal, S. K. 1978. *The computed tomography of brain abscesses.* Presented at 16th Ann. Meet. Am. Soc. Neuroradiol., New Orleans
17. Boltshauser, E., Hamalatha, H., Grant, D. N., Tiel, K. 1977. Impact of computerized axial tomography on the management of posterior fossa tumors in childhood. *J. Neurol. Neurosurg. Psychiatry* 40:209-13
18. Larson, E. B., Omenn, G. S., Loop, J. W. 1978. Computed tomography in patients with cerebrovascular disease: impact of a new technology in patient care. *Am. J. Roentgenol.* 131:35-40
19. Abrams, H. L., McNeil, B. J. 1978. Medical implications of computed tomography I. *N. Engl. J. Med.* 298: 255-61
20. Abrams, H. L., McNeil, B. J. 1978. Medical implications of computed tomography II. *N. Engl. J. Med.* 298: 310-18
21. Abrams, H. L., McNeil, B. J. 1978. Computed tomography: cost and efficacy implications. *Am. J. Roentgenol.* 131:81-87
22. Backman, D. S., Hodges, F. J. III, Freeman, J. M. 1976. Computed axial tomography in chronic seizure disorders of childhood. *Pediatrics* 58:828-32
23. Kingsley, D., Kendall, B. E. 1978. The value of computed tomography in the evaluation of the enlarged head. *Neuroradiology* 15:59-71
24. Knaus, W. A., Davis, D. O. 1978. Utilization and cost-effectiveness of cranial computed tomography at a university hospital. *J. Comput. Assist. Tomogr.* 2:209-14
25. Larson, E. B., Omenn, G. S., Margulis, M. T., Loop, J. W. 1977. Impact of computed tomography on utilization of cerebral angiograms. *Am. J. Roentgenol.* 129:1-3
26. Bahr, A. L., Hodges, F. J. III. 1978. Efficacy of computerized tomography of the head in changing patient care and health cost: A retrospective study. *Am. J. Roentgenol.* 131:45-49
27. Davis, K. R., Taveras, J. M., Roberson, G. H., Ackerman, R. H. 1976. Some limitations of computed tomography in

- the diagnosis of neurological diseases. *Am. J. Roentgenol.* 127:111-23
28. Wing, S. D., Osborn, A. G., Wing, R. W. 1978. The vertex scan: An important component of cranial computed tomography. *Am. J. Roentgenol.* 130: 765-67
29. Rosenbaum, H. E. 1977. Computerized tomographic scans of brain: Three-dimensional. *Arch. Neurol.* 34:386-87
30. Osborn, A. G., Anderson, R. E. 1978. Direct sagittal CT scans of the face and paranasal sinuses. *Radiology.* 129: 81-87
31. Bergström, M., Greitz, T. 1976. Stéréotaxic computed tomography. *Am. J. Roentgenol.* 127:167-70
32. Tadmor, R., Davis, K. R., Roberson, G. H., New, P. F. J., Taveras, J. M. 1978. Computed tomographic evaluation of traumatic spinal injuries. *Radiology* 127:825-27
33. Wolpert, S. M., Scott, R. M., Carter, B. L. 1977. Computed tomography in spinal dysraphism. *Surg. Neurol.* 8:199-206
34. Drayer, B. P., Rosenbaum, A. E., Hyman, H. B. 1977. Cerebrospinal fluid imaging using serial metrizamide CT cisternography. *Neuroradiology* 13: 7-17
35. Phelps, M. E., Hoffman, E. J., Huang, S. C., Kuhl, D. E. 1978. A new computerized tomographic imaging system for positron-emitting radiopharmaceuticals. *J. Nucl. Med.* 19:635-47

LOW MOLECULAR WEIGHT PROTEINURIA

♦7316

C. L. Hall, M.D., M.R.C.P.

Renal Research Laboratory, Queen Elizabeth Hospital,
Birmingham B15 2TH, England

J. Hardwicke, M.D., M.R.C.P., F.R.C. Path.

Department of Immunology, University of Birmingham,
Birmingham B15 2TJ, England

INTRODUCTION

Low molecular weight (LMW) proteinuria is defined in this article as the excretion in the urine of proteins of a size smaller than albumin [MW 69,000; molecular radius (MR) 37Å]. The LMW proteins may enter the urine either from the plasma by filtration through the normal glomerular basement membrane (GBM) or by secretion or loss from the nephron. However, most of the proteins secreted by the nephron that have been identified to date are of high molecular weight (HMW) e.g. Tamm Horsfall protein, secretory IgA. At present, therefore, the LMW proteins found in the urine seem to originate by the filtration of LMW plasma proteins at the glomerulus. The LMW proteins present normally include enzymes (lysozyme, amylase), polypeptide hormones (insulin and growth hormone), polyclonal immunoglobulin fragments, fibrin-fibrinogen degradation products, and LMW proteins of unknown function (α_2 and β_2 microglobulin, post- γ proteins). In some disease states LMW proteins not normally present in the plasma (myoglobin, hemoglobin, monoclonal immunoglobulin fragments) may appear and may enter the urine, and thus provide us with valuable diagnostic markers.

RENAL HANDLING OF LOW MOLECULAR WEIGHT PROTEINS

Glomerular Permeability

Dextran and polyvinyl pyrrolidone, inert polymers of defined sizes, have been used to investigate the glomerular permeability to LMW proteins (1-4). The glomerular clearance is inversely related to the size of the polymer; it is 95% of the glomerular filtration rate (GFR) for small particles of 25Å and 5% of GFR for large particles of 50Å. This conclusion is supported by studies of the clearance of individual proteins by the glomerulus (Table 1). However, some LMW proteins behave in an anomalous manner. For example, the glomerular clearance of lysozyme is less than expected for its molecular size when isolated in pure form (7). This is probably because monomeric lysozyme is in equilibrium with a dimeric form at physiological pH in the plasma (8). A similar explanation may account for the absence of the LMW plasma proteins orosomucoid (MW 40,000) and α_1 antitrypsin (MW 45,000) from normal urine (9).

Tubular Reabsorptive Mechanisms

Use of purified LMW proteins (6, 10-13) demonstrated that, after filtration at the glomerulus, the proteins are reabsorbed by the renal tubules. A recent study employing ^{125}I -labeled lysozyme and an isolated perfused rat kidney system demonstrated that the entire plasma lysozyme pool was filtered every hour while less than 0.5% of the filtered load was excreted in the urine (6). The reabsorption of lysozyme was an active, energy dependent function of the proximal tubular cells and a saturable process with a maximum capacity in excess of the usual filtered load. Increased urinary lysozyme excretion began at filtered loads 1-2 times above normal, increased linearly with larger filtered loads (below saturation), and then paralleled the filtered load when tubular reabsorptive mechanisms were saturated. Electron microscopy and autoradiography (14-16) reveal that, after injection, LMW proteins rapidly become associated with the brush borders of the proximal

Table 1 Glomerular clearance of LMW proteins in relation to molecular size

Protein	Molecular weight (daltons)	Molecular radius (Å)	Clearance (% GFR)	Reference
Albumin	69,000	37	0.02	3
Amylase	48,000	29	3	5
Light-chain dimer	44,000	27	8	6
Myoglobin	16,000	20	75	7
Lysozyme	15,000	19	80	8

tubular cells in small apical vesicles; these subsequently form phagosomes, which fuse with lysosomes to form intracytoplasmic phagolysosome bodies. LMW proteins may also be absorbed directly into the cytosol, which under normal conditions may be the main reabsorptive pathway (14), with reabsorption into phagolysosome bodies occurring at higher filtered loads.

Tubular Transport and Catabolism

Much experimental work indicates that reabsorbed LMW proteins are catabolized by the enzymes of the proximal tubular cells to peptides and amino acids (17, 18). However, studies of rat kidneys loaded in vivo with radiolabeled lysozyme or insulin and then maintained by perfusion in vitro show that both intact proteins and catabolic products are released into the perfusate. Some 40% of the lysozyme was released as intact protein and 60% as catabolic products, whereas more than 90% of the insulin was released as catabolic products (6, 14). The mechanisms whereby some proteins are returned intact and others totally catabolized are not understood but may be related to differences in molecular size, susceptibility to proteolytic enzymes, or tubular transport pathways.

Tubular Contraluminal Transport

Radiolabeled insulin (19) and lysozyme (6) were used in an isolated perfused renal tubular system to demonstrate that the contraluminal uptake of these proteins is less than 3.5% of the luminal uptake. Thus despite early reports (12) there is little firm evidence to support a significant contraluminal uptake of LMW protein by the renal tubular cells.

Independent Tubular Reabsorptive Mechanisms

The tubular reabsorption of proteins of albumin size (MW 69,000; MR 37 Å) or greater appears to be a nonselective process; the proteins are reabsorbed in proportion to their concentration in the tubular luminal fluid (20). There is considerable evidence that HMW and LMW proteins are reabsorbed by different tubular mechanisms (7, 17). For example in the nephrotic syndrome the urinary clearance and metabolism of LMW proteins remains within normal limits despite glomerular filtration and urinary loss of large amounts of HMW proteins (5, 21). In contrast, conflicting data exist concerning the possible competition between individual LMW proteins for tubular reabsorption. In dogs the intravenous infusion of Bence Jones protein, myoglobin, or ovalbumin did not alter the urinary excretion of endogenous or infused lysozyme (22). However, in rats an injection of ovalbumin significantly decreased the renal tubular reabsorption of horseradish peroxidase (23). Furthermore the urine of a patient with paroxysmal myoglobinuria contained increased amounts of several LMW proteins in

addition to myoglobin (24). Thus it appears that both specific and competitive tubular reabsorptive pathways may exist for some LMW proteins and that these may differ in different species.

CLINICAL TYPES OF PROTEINURIA

Normal Proteinuria

The normal urinary protein excretion of healthy people varies between 24 and 133 mg per day (9). Approximately half of this consists of the many LMW plasma proteins found in greater amounts in tubular proteinuria. The remainder consists mainly of albumin and larger plasma proteins unselectively lost into the glomerular filtrate or from the interstitial fluid of the kidney, together with proteins secreted by the nephron (e.g. Tamm Horsfall glycoprotein, secretory IgA) and, in the male, proteins secreted by the prostate and seminal vesicles (25).

Glomerular Proteinuria

In glomerular proteinuria (Figure 1) proteins the size of albumin or greater are lost in the urine through an abnormally permeable GBM (26). The urinary albumin clearance exceeds 0.02% of the GFR while the clearance of LMW proteins remains within normal limits and quantitatively negligible (3, 21). The total urinary protein excretion in glomerular disease may vary between as little as 1 gram and as much as 30 grams daily and usually is considerably greater than in tubular disease (3).

Tubular Proteinuria

In tubular proteinuria (Figure 2) the urinary albumin clearance remains within normal limits (3) and albumin comprises only 5–25% of the urinary protein (5, 27). Most of the urinary protein consists of the 15–50 different LMW proteins (MW 10,000–50,000) filtered from the plasma through the normal glomeruli and not reabsorbed by the malfunctioning renal tubules (28–30). Thus the major difference between tubular and normal proteinuria is quantitative. In tubular proteinuria the LMW proteins comprise more than 50% of the urinary protein while albumin constitutes less than 25%, and the total urinary protein excretion exceeds 150 mg per day but rarely exceeds 2 g per day (3, 30). Thus the detection in the urine of several LMW proteins, such as α_2 and β_2 microglobulin, lysozyme, post- γ -globulin, and light chains, in amounts exceeding 150 mg per day, whether or not accompanied by HMW proteins, indicates the failure of tubular reabsorption provided that the plasma levels of the LMW proteins are within normal limits. There are two main situations described in which increased plasma levels of LMW proteins may be found.

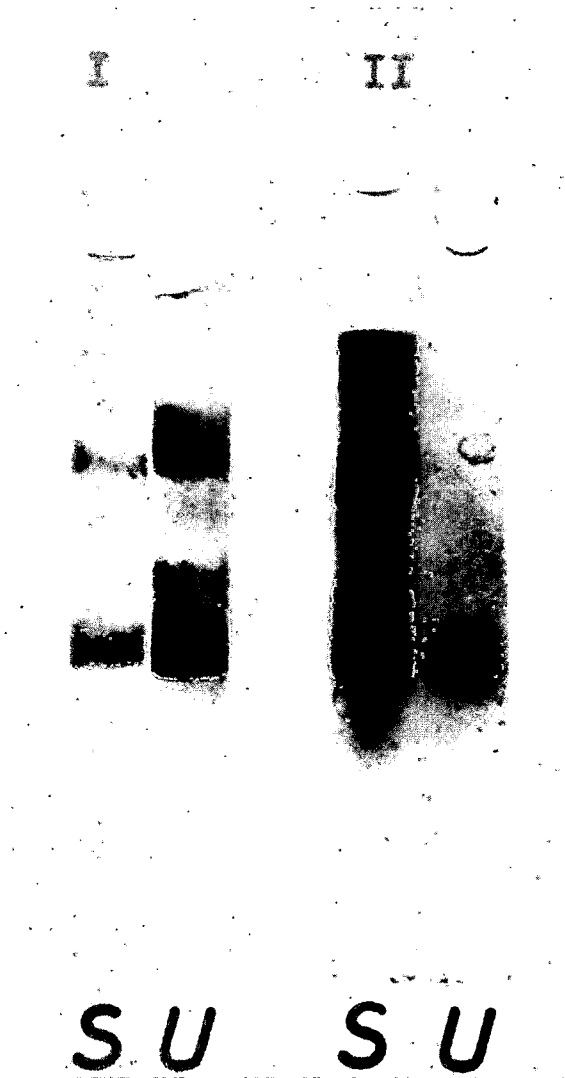


Figure 1 SDS polyacrylamide gel electrophoresis of serum (S) and urine (U) of two patients with glomerular proteinuria. (I) Nonselective proteinuria—both albumin and high molecular weight globulins detected in the urine. (II) Selective proteinuria—only albumin detected in the urine.



Figure 2 SDS polyacrylamide gel electrophoresis of serum (S) and urine (U) of a patient with tubular proteinuria showing several LMW proteins as well as albumin in the urine.

GENERALIZED INCREASE IN PLASMA LMW PROTEIN In chronic renal impairment (creatinine clearance < 10 ml/min) the plasma levels of many LMW proteins are increased, since renal tubular catabolism of these proteins is decreased (5). The increased filtered load of LMW proteins per surviving nephron exceeds the tubular reabsorptive capacity and increased quantities of LMW proteins enter the urine. Thus glomerular and tubular proteinuria may occur together in any glomerular disease that has progressed to chronic renal impairment (20), as well as in diseases that damage both the glomeruli and tubules, e.g. acute renal failure, kidney allograft rejection, and chronic pyelonephritis (31).

INCREASED PLASMA LEVEL OF ONE LMW PROTEIN Increased levels of a specific LMW protein may occur in the plasma (e.g. amylase, lysozyme, monoclonal light or heavy chains, myoglobin, hemoglobin). In these circumstances the filtered load of the protein may exceed the tubular reabsorptive capacity, which will lead to its excretion in the urine in large amounts

(4). However, as reabsorptive mechanisms for LMW proteins are, at least in part, competitive and nonspecific, their saturation by one LMW protein leads to the urinary loss of many LMW proteins, although the component with the increased plasma concentration predominates. This has been observed clinically in a patient with paroxysmal myoglobinuria whose urine contained increased levels of retinol-binding protein and β_2 microglobulin in addition to myoglobin (24). The clinical significance of increased serum and urinary levels of individual LMW proteins is discussed below.

Thus tubular proteinuria is defined by the presence of increased amounts of several LMW proteins in the urine in the absence of increased plasma levels. Tubular proteinuria is found usually in patients with histological evidence of tubular and/or interstitial renal disease and/or biochemical evidence of tubular dysfunction, although it may occur occasionally as an isolated abnormality, e.g. familial asymptomatic tubular proteinuria (31). A classification of the conditions in which tubular proteinuria has been detected is presented in Table 2.

In some conditions, e.g. cadmium nephropathy and hereditary and acquired Fanconi syndrome (28, 32–34), tubular proteinuria alone is usually detected. However, in other conditions, e.g. pyelonephritis, acute renal failure, kidney allograft rejection, and Balkan nephropathy (31, 33), some patients have tubular proteinuria, others glomerular proteinuria, and many a combination of both types.

Specific tubular diseases may be associated with characteristic patterns of LMW proteinuria (31). However, patients with the same disease may excrete different LMW proteins in varying concentrations and individual

Table 2 Causes of low molecular weight proteinuria

Congenital tubulopathy	Fanconi syndrome, oculocerebrorenal dystrophy, cystinosis, renal tubular acidosis, Bartter's syndrome, nephrogenic diabetes insipidus, familial asymptomatic tubular proteinuria
Systemic disease	
Hereditary	Hepatolenticular degeneration, galactosemia, glycogen storage disease, cystinosis
Acquired	Sarcoidosis, multiple myeloma, Balkan nephropathy
Drugs	Drug-induced interstitial nephritis, e.g., methicillin, laxative abuse
Heavy metals and poisons	Cadmium, lead, arsenic, mercury, ethylene glycol, carbon tetrachloride
Acute renal disease	Acute tubular necrosis, kidney allograft rejection
Infective	Acute/chronic pyelonephritis
Miscellaneous	Nephrocalcinosis, nephronophthisis, proteinuria of hypothermia

patients may excrete varying amounts of different proteins at different times (27). Thus, with the exception of conditions associated with the abnormal production of a single LMW protein and its excretion in the urine (see below), specific patterns of LMW proteinuria corresponding with individual diseases have not been identified.

SIGNIFICANCE OF THE EXCRETION OF PREDOMINANTLY ONE LMW PROTEIN

Amylase

The clearance of amylase (MW 45,000–50,000) is normally about 3% of the creatinine clearance (4, 35), but it increases in the presence of tubular damage, uremia, and increased glomerular permeability (35). In acute pancreatitis the urinary amylase clearance increases significantly, often to more than 10% of the creatinine clearance (36). Initially it was thought that this was specific for acute pancreatitis because it was not observed in patients with raised serum amylase levels resulting from other conditions, and thus it was ascribed to altered renal permeability to amylase in acute pancreatitis (36). Subsequently, however, increased urinary amylase clearance has been observed in patients with burns, diabetic ketoacidosis, perforated peptic ulcer, and postsurgical or accidental trauma (37, 38) associated with increased urinary excretion of other LMW proteins (38, 39). As the patients studied did not have significant depression of the GFR, these data indicate that the increased amylase clearance is due either to tubular damage caused by hypotension, hyperosmolar plasma, proteolytic enzymes, and myoglobinuria (40) or to saturation of the tubular reabsorptive sites by large amounts of protein catabolic or degradation products (41).

Light Chains

Kappa (κ) and lambda (λ) light (L) chains are synthesized by lymphocytes and plasma cells. Small amounts of polyclonal L chains are present in normal urine (20) and in considerably increased amounts in the urine of patients with renal tubular dysfunction (5, 42). Large amounts of L chain are found in the urine of some patients with proliferative disorders of lymphocytes and plasma cells such as myeloma, macroglobulinemia, and lymphatic leukemia, when the L chains are monoclonal (43). The clearance of L chains by the glomerulus varies between 7 and 45% (44) of creatinine clearance according to the ratio of L-chain dimer (MW 44,000) to monomer (MW 22,000) present in the plasma.

The presence of large amounts of L chain in the urine, which precipitate at 55–60°C and redissolve at 95°C, are the cause of Bence Jones proteinuria (BJP). Not all patients with myeloma have BJP (43). Some have only HMW monoclonal protein, which does not pass the normal GBM. However,

HMW monoclonal proteins may appear in the urine as a result of increased GBM permeability associated with the glomerular damage that may occur in myeloma (45). Renal failure is common in patients with myeloma and is frequently caused by amyloid, hypercalcemia, hyperuricemia, and pyelonephritis (2). Occasionally an acquired Fanconi syndrome develops possibly owing to the nephrotoxic effects of L chains reabsorbed by the tubular cells (34). In vitro experiments with rabbit and rat cortical slices demonstrated that tubular transport processes are impaired by impure myeloma protein (46), although this has not been confirmed using purified BJP (47), and it is unresolved as to whether both κ and λ L chains are equally toxic to tubular cells (20). Increased serum levels of polyclonal L chains occur only rarely, usually in association with chronic renal impairment (43), and are the result of decreased glomerular filtration and decreased tubular catabolism (5). However, increased serum monoclonal L-chain levels occur occasionally without increased urinary levels and are due to the formation of L-chain polymers too large for glomerular filtration (20). Occasionally L-chain fragments, corresponding to the variable portion of the molecule, are excreted in the urine and represent partially degraded or incompletely synthesized L chain (42).

Fc Fragment and Heavy Chains

Because of their considerable molecular size, the clearance of Fc fragments (MW 47,000) and heavy (H) chains (MW 55,000) by the glomerulus is less than 5% of the clearance of creatinine. Furthermore H chains in serum and plasma tend to polymerize and become too large for glomerular filtration (48). Normal urine contains about 0.19 mg of polyclonal Fc fragments per day, which are derived from the endogenous catabolism of IgG and comprise up to 15% of the urinary λ -globulin (49). Unlike L chains, Fc fragment does not appear to be catabolized significantly by the renal tubules (11) and increased urinary excretion of monoclonal Fc fragment has been observed in some 3% of patients with multiple myeloma (50). H chains may be present in the urine of patients with proliferative disorders of plasma cells and lymphocytes (48) and are independent monoclonal synthetic products rather than immunoglobulin degradation products (51). Urinary excretion of gamma and alpha H chains has been reported in patients with IgG myeloma and with malignant intestinal lymphoma, respectively, and mu H chains may be found in patients with macroglobulinemia and chronic lymphatic leukemia (48).

Fibrin-Fibrinogen Degradation Products

Normal urine contains only small amounts of the LMW fibrin-fibrinogen degradation products (FDP) fragments D and E (52). In dogs infused with purified FDP fragments (53) and in patients receiving thrombolytic therapy

(54) only small amounts of the LMW fragments appeared in the urine. Renal tubular reabsorption of the LMW fragments D and E has been demonstrated by immunofluorescent techniques in rats (55). The failure of tubular reabsorptive mechanisms in renal tubular and interstitial disease is associated with a small but significant increase in urinary FDP excretion (56, 57). Large increases in the urinary FDP excretion ($> 10 \mu\text{g/ml}$), which may occur in kidney allograft rejection and in some patients with acute nephritis and the nephrotic syndrome, occur only in association with increased glomerular permeability, and the FDP consist mainly of the HMW fragments X and Y and intact fibrinogen (58, 59) filtered from the plasma.

Myoglobin and Hemoglobin

Myoglobin is the LMW (16,000) oxygen-binding protein of muscle, which when released into the circulation is not bound to carrier proteins (60). It has a low renal threshold (0.3 mg%), and urine containing myoglobin exhibits a brownish discoloration, gives positive reaction with benzidine and orthotolidine reagents, and frequently contains pigmented granular casts (61, 62). Myoglobin release into the plasma occurs in association with rhabdomyolysis, which also gives rise to raised serum levels of muscle enzymes, (creatine phosphokinase and aldolase). It may be due either to hereditary disorders of muscle metabolism (e.g. McArdle's syndrome, primary paroxysmal myoglobinuria) or to muscle damage associated with ischemia, surgical or accidental trauma, drugs, etc (61–63). Myoglobinuria is associated with increased urinary excretion of other LMW proteins either because the myoglobin causes tubular damage or because other LMW proteins are competing for tubular reabsorptive sites (24).

Hemoglobinuria occurs in association with intravascular hemolysis or hematuria. Hematuria is identified by the presence of red cells and casts in the urine, while intravascular hemolysis is associated with a pink discoloration of the separated serum and reduced levels of serum haptoglobins (61). The glomerular clearance of hemoglobin (MW 68,000) is low and the renal threshold high (128 mg%), since the binding of hemoglobin to plasma haptoglobin forms a complex too large for glomerular filtration (64). It is only when the binding capacity of haptoglobin is exceeded that hemoglobin appears in the urine.

In both man and animals the infusion of hemoglobin or myoglobin seldom causes significant impairment of renal function (62). However, in the presence of shock, hypotension, dehydration, and acidosis, all of which frequently accompany the conditions associated with hemoglobin and myoglobin release into the plasma, acute renal failure often occurs and appears to be at least in part a result of tubular obstruction by hemoglobin- and myoglobin-containing casts.

CONCLUSION

LMW proteinuria is a clinical entity, pathologically distinct from "albuminuria." The same LMW proteins that are present in normal urine are found in increased amounts in pathological LMW proteinuria. Each individual LMW protein is filtered from the plasma through the normal GBM and has a glomerular clearance determined by its molecular size. The proteins are reabsorbed by the renal tubules so that normal urine contains less than 50–60 mg of LMW proteins per day. In diseases that affect the renal tubules the reabsorptive mechanisms become impaired and increased amounts of the many different LMW proteins are excreted in the urine. This occurs in the absence of increased plasma levels except in advanced renal impairment when the GFR is reduced sufficiently to prevent the plasma clearance and tubular catabolism of LMW protein.

Pathological LMW proteinuria may also be caused by the excretion of a particular protein. This occurs when increased amounts of the protein are being synthesized in the body and filtered at the glomerulus, exceeding tubular reabsorptive capacity. Due to competition between the LMW proteins for tubular reabsorptive sites, the presence of one LMW protein in the urine may be accompanied by several others, although at lower concentration. If the GFR is normal or only moderately reduced, the serum level of the LMW protein may (hemoglobin, amylase) or may not (L chains) be increased, but the serum level rises if advanced renal impairment develops.

Thus the detection in the urine of increased levels of several LMW proteins indicates disease affecting the renal tubules, whereas the detection of predominantly one species of LMW protein may indicate disease affecting other organs or systems, e.g. myeloma, pancreatitis, rhabdomyolysis, intravascular hemolysis, or fibrinolysis. Loss of LMW protein into the urine as the result of tubular secretion or parenchymal damage has yet to be shown to be of clinical significance.

Literature Cited

1. Hardwicke, J., Hulme, B., Jones, J. H., Ricketts, C. R. 1968. Measurement of glomerular permeability to polydisperse radioactively labelled macromolecules in normal rabbits. *Clin. Sci.* 34:505–14
2. Arturson, G., Groth, T., Grotte, G. 1971. Human glomerular membrane porosity and filtration pressure. *Clin. Sci.* 40:137–58
3. Hardwicke, J. 1975. Laboratory aspects of proteinuria in human disease. *Clin. Nephrol.* 3:37–41
4. Blainey, J. D., Northam, B. E. 1967. Amylase excretion by the human kidney. *Clin. Sci.* 32:377–83
5. Waldmann, T. A., Strober, W., Mogielnicki, R. P. 1972. Renal handling of low molecular weight proteins. *J. Clin. Invest.* 51:2162–74
6. Maack, T. 1975. Renal handling of low molecular weight proteins. *Am. J. Med.* 58:57–64
7. Harrison, J. F., Lunt, G. S., Scott, P., Blainey, J. D. 1968. Urinary lysozyme ribonuclease and low molecular weight protein in renal disease. *Lancet* 1:371–75
8. Sophianopoulos, A. J., van Holde, K. E. 1964. Physical studies of lysozyme. *J. Biol. Chem.* 239:2516–24

9. Berggard, I. 1970. Plasma proteins in normal human urine. In *Proteins in Normal and Pathological Urine*, ed. Y. Manuel, J. P. Revillard, H. Betuel. Basel: Karger. pp. 7-19
10. Bernier, G. M., Conrad, M. E. 1969. Catabolism of β_2 microglobulin in the rat kidney. *Am. J. Physiol.* 217:1359-62
11. Wochner, R. D., Strober, W., Waldmann, T. A. 1967. The role of the kidney in the catabolism of Bence Jones proteins and immunoglobulin fragments. *J. Exp. Med.* 126:207-21
12. Chamberlain, M. J., Stimmler, L. 1967. The renal handling of insulin. *J. Clin. Invest.* 46:911-19
13. Katz, A. I., Rubenstein, A. H. 1973. Metabolism of proinsulin, insulin and C-peptide in the rat. *J. Clin. Invest.* 52:1113-21
14. Maack, T., MacKensie, D. D., Kinter, W. B. 1971. Intracellular pathways and renal reabsorption of lysozyme. *Am. J. Pathol.* 221:1609-16
15. Oliver, C., Essner, E. 1972. Protein transport in mouse kidney utilizing tyrosinase as an ultrastructural tracer. *J. Exp. Med.* 136:291-304
16. Bourdeau, J. E., Carone, F. A. 1974. Protein handling by the renal tubules. *Nephron* 13:22-34
17. Maack, T. 1967. Changes in activity of acid hydrolase during the renal reabsorption of lysozyme. *J. Cell Biol.* 35:268-73
18. Izzo, J. L., Racone, A., Izzo, M. J. 1971. Degradation of 131 I proinsulin and 131 I insulin by rat liver, kidney and muscle extracts. *Fed. Proc.* 30:193
19. Rabkin, R., Rubenstein, A. H., Colwell, J. A. 1972. Glomerular filtration and proximal tubular reabsorption of 125 I insulin. *Am. J. Physiol.* 223:1093-96
20. Hardwicke, J., Cameron, J. S., Harrison, J. F., Hulme, B., Soothill, J. 1970. See Ref. 9, pp. 111-52
21. Boesken, W. H., Kopf, K., Schollmeyer, P. 1973. Differentiation of proteinuric diseases by disc electrophoretic molecular weight analysis of urinary proteins. *Clin. Nephrol.* 1:111-18
22. Harrison, J. F., Barnes, A. D. 1970. The urinary excretion of lysozyme in dogs. *Clin. Sci.* 38:533-47
23. Straus, W. 1962. Colorimetric investigation of uptake of horseradish peroxidase by rat kidney and effects of competition by egg white. *J. Cell. Biol.* 12:231-46
24. Ravnskov, U. 1975. Low molecular weight proteinuria in association with paroxysmal myoglobinuria. *Clin. Nephrol.* 3:65-69
25. Blainey, J. D., Brewer, D. B., Hardwicke, J. 1978. Proteinuria and the nephrotic syndrome. In *Renal Disease*, ed. D. A. K. Black, Oxford: Blackwell. 3rd ed. In press
26. Blainey, J. D., Brewer, D. B., Hardwicke, J., Soothill, J. F. 1960. The nephrotic syndrome: diagnosis by renal biopsy and biochemical and immunological analyses related to the response to steroid therapy. *Q. J. Med.* 29:235-56
27. Manuel, Y., Laterra, E. C. 1970. See Ref. 9, pp. 172-87
28. Piscator, M. 1966. Proteinuria in chronic cadmium poisoning. *Arch. Environ. Health* 12:335-44
29. Walravens, P., Laterra, E. C., Heremans, J. F. 1968. Studies on tubular proteinuria. *Clin. Chim. Acta* 19:107-20
30. Flynn, F. V., Platt, H. S. 1968. The origin of the proteins excreted in tubular proteinuria. *Clin. Chim. Acta* 21:377-99
31. Revillard, J. P., Manuel, Y., Francois, R., Traeger, T. 1970. See Ref. 9, pp. 209-19
32. Creeth, J. M., Kekwick, R. A., Flynn, F. V., Harris, H., Robson, E. B. 1963. An ultracentrifuge study of proteinuria in renal tubular disorders. *Clin. Chim. Acta* 8:406-13
33. Davis, J. S., Flynn, F. V., Platt, H. S. 1968. Characterization of urinary proteins by gel filtration. *Clin. Chim. Acta* 21:357-76
34. Harrison, J. F., Blainey, J. D. 1967. Low molecular weight proteinuria in chronic renal disease. *Clin. Sci.* 33:381-90
35. Fredricksson, A., Jacobson, G. 1974. Influence of renal function on the urinary excretion of amylase in the rat. *Scand. J. Lab. Clin. Invest.* 33:313-21
36. Warshaw, A. L., Fuller, A. F. 1975. Specificity of increased renal clearance of amylase in the diagnosis of acute pancreatitis. *N. Engl. J. Med.* 292:325-28
37. Levine, R. I., Glauser, F. L., Berk, J. E. 1975. Enhancement of amylase, creatinine clearance ratio in disorders other than acute pancreatitis. *N. Engl. J. Med.* 292:329-32
38. Berger, G. M., Cowlin, J., Turner, T. J. 1976. Amylase-creatinine clearance ratio and urinary excretion of lysozyme in acute pancreatitis and acute duodenal perforation. *S. Afr. Med. J.* 50:1559-61

39. Wide, L., Thoren, L. 1972. Increased urinary excretion of albumin, β_2 , microglobulin, insulin and luteinizing hormone following surgical or accidental trauma. *Scand. J. Clin. Lab. Invest.* 30: 275-81
40. Weeden, R. R., Samuels, A. J. 1975. Amylase-creatinine clearance ratio in pancreatitis. *N. Engl. J. Med.* 292:923
41. Jacobson, G. 1975. Amylase-creatinine clearance ratio in pancreatitis. *N. Engl. J. Med.* 292:923
42. Creysell, R. 1970. See Ref. 9, pp. 100-11
43. Bauxbaum, J. N., Franklin, E. C. 1970. See Ref. 9, pp. 83-91
44. Harrison, J. F., Blainey, J. D., Hardwicke, J., Rowe, D. S., Soothill, J. F. 1966. Proteinuria in multiple myeloma. *Clin. Sci.* 31:95-110
45. Abrahams, C., Pirani, C. L., Pollack, V. E. 1966. Ultrastructure of the kidney in a patient with multiple myeloma. *J. Pathol. Bacteriol.* 92:220-24
46. Preus, H. G., Weiss, F. R., Iammarino, R. M., Hammack, W. J., Murdaugh, H. V. 1974. Effects of rat kidney slice function in vitro of proteins from the urines of patients with myelomatosis and nephrosis. *Clin. Sci. Mol. Med.* 46:283-94
47. Fine, J. D., Rees, E. 1977. Renal cytotoxic effects of Bence Jones proteinuria. *Lancet* 1:150
48. Franklin, E. C. 1970. Heavy chain diseases. *N. Engl. J. Med.* 282:1098-99
49. Vaughan, J. H., Ralph, F. J., Gray, B. A. 1967. Light and heavy chain components of γ globulin in urines of normal persons and patients with agammaglobulinaemia. *J. Clin. Invest.* 46: 266-79
50. Solling, K., Askjaer, S. A. 1973. Multiple myeloma and urinary excretion of heavy chain components of IgG. *Acta Med. Scand.* 194:22-30
51. Seligman, M., Mihaesco, E., Hurez, D., Mihaesco, C., Preud'Homme, J.-L., Rambud, J. C. 1969. Immunochemical studies in four cases of alpha chain disease. *J. Clin. Invest.* 48:2374-89
52. Poortmans, J., Luke, K., Zipursky, A., Bienenstock, J. 1971. Fibrin split products in exercise proteinuria. *Clin. Chim. Acta* 35:449-54
53. Raynor, H., Paraskervas, F., Israels, L. G., Israels, E. D. 1969. Fibrin degradation products; identification and assay in serum and urine. *J. Lab. Clin. Med.* 74:586-96
54. Bouma, B., Hedner, U., Nilsson, I. 1971. Typing of fibrinogen degradation products in various clinical disorders. *Scand. J. Lab. Clin. Invest.* 27:331-35
55. Salmon, J., Lambert, P. H. 1971. Determination of fibrinogen derivatives by immunofluorescence in experimental and clinical kidney disorders. *Scand. J. Haematol.* 13: Suppl., pp. 351-56
56. Ekert, H., Barrett, M., Chantler, C., Turner, M. 1972. Immunological related equivalents of fibrinogen in the sera and urine of children with renal disease. *Arch. Dis. Child.* 47:90-96
57. Naish, P., Evans, D., Peters, D. 1974. Urinary FDP excretion and intraglomerular fibrin deposits in glomerulonephritis. *Br. Med. J.* 1:544-47
58. Hall, C. L., Pejhan, N., Terry, J. M., Blainey, J. D. 1975. Urinary fibrin-fibrinogen degradation products in the nephrotic syndrome. *Br. Med. J.* 1: 419-22
59. Hall, C. L., Blainey, J. D., Gaffney, P. J. 1978. The origin of urinary fibrin-fibrinogen degradation products in renal glomerular disease. *Nephron*. In press
60. Markowitz, H., Wobig, G. H. 1977. Quantitative method of estimating myoglobin in the urine. *Clin. Chem.* 23: 1689-93
61. Rowland, L. P., Penn, A. S. 1972. Myoglobinuria. *Med. Clin. North Am.* 56:1233-45
62. Grossman, R. A., Hamilton, R. W., Morse, B. M., Penn, A. S., Goldberg, M. 1974. Non traumatic rhabdomyolysis and acute renal failure. *N. Engl. J. Med.* 291:807-11
63. Koffler, A., Friedler, R. M., Massery, S. G. 1976. Acute renal failure due to non-traumatic rhabdomyolysis. *Ann. Intern. Med.* 85:22-28
64. Allison, A. C., Rees, W. 1957. The binding of haemoglobin by plasma proteins. *Br. Med. J.* 2:1137-43

THE ULTRASTRUCTURE OF GLOMERULAR FILTRATION

♦7317

Morris J. Karnovsky, M.D.

Department of Pathology, Harvard Medical School,
Boston, Massachusetts 02115

The glomeruli of the kidney serve to separate about one third of the plasma entering them into a virtually ideal ultrafiltrate. To accomplish this, fluid is driven by hydraulic forces across the glomerular capillary wall (GCW), which has a high permeability to water but virtually excludes molecules larger than albumin from the filtrate.

In recent years it has become apparent that the glomerulus is not merely a mechanical filter but that it is an organ of extreme complexity and subtle behavior. It is the purpose of this brief review to cover the physiological and morphological observations underlying our current concepts of the structural bases for glomerular filtration.

ULTRASTRUCTURE OF THE GLOMERULUS

The anatomic structures in the mammalian glomerulus that could act as barriers to the passage of macromolecules are: (a) the endothelium; (b) the glomerular basement membrane; and (c) the epithelial layer. For a detailed review on glomerular structure see Reference 1. In the discussion below refer to Figures 1 and 2.

The Endothelium

The endothelial layer consists of attenuated endothelial cells spread along the glomerular basement membrane. The cells are about 40 nm in thickness for most of their length, and their most notable characteristic is that the thin cell bodies are completely penetrated by open holes or fenestrae ("windows"). These fenestrae are virtually round and vary in diameter from 40-100 nm.



Figure 1. Low power electron micrograph of a portion of a glomerulus, fixed by perfusion. L = capillary lumen; E = endothelium (cell body containing nucleus and attenuated fenestrated portion); X = *en face* view of fenestrated portion of a capillary: note the holes (fenestrae) penetrating the cytoplasm; EC = epithelial cells, extending foot processes (FP) to the glomerular basement membrane: the middle layer (lamina densa) of the basement membrane appears as a continuous dark line; US = urinary space; M = mesangial cells interposed between the capillary loops; rectangle = area diagrammed in Figure 2. Enlarged 3750 times.

The Glomerular Basement Membrane

The glomerular basement membrane (GBM) as visualized from suitable heavy metal staining, consists of three layers: (a) the lamina rara interna, a subendothelial or inner layer that is relatively electron lucent; (b) the lamina densa, a middle or central layer that is electron opaque and finely fibrillar; and (c) the lamina rara externa, an outer, electron-lucent subepithelial layer. In man, the width of the GBM is 240–340 nm, and each layer occupies approximately one third of the total width.

CHEMISTRY OF THE GBM Chemical analysis of the GBM (2–4) reveals two distinct classes of glycopeptides. The first class consists of a relatively nonpolar collagen-like component containing large amounts of hydroxyproline, glycine, and hydroxylysine. This fraction has a high hexose content, with a large number of glucosyl-galactose disaccharide units in glycosidic linkage to hydroxylysine; there are few heteropolysaccharide units. The

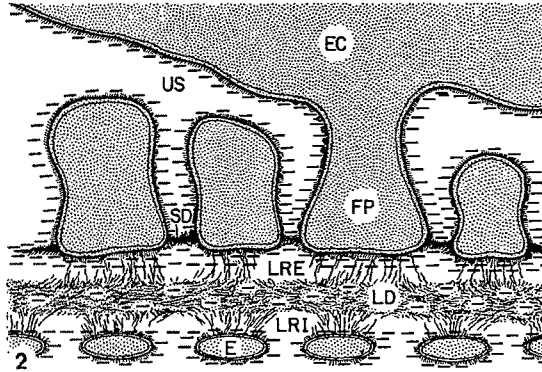


Figure 2. Diagram of the glomerular filtration barrier at high magnification of an area similar to that seen in the rectangle in Figure 1. E = endothelium perforated by fenestrae; LRI = lamina rara interna; LD = lamina densa; LRE = lamina rara externa; FP = foot process; SD = slit diaphragm extending between adjacent foot processes; EC = epithelial cell; US = urinary space. The distribution of fixed negative charge is indicated by dashes. Note the fibrillar nature of the LD. Mesangial cells are not depicted. Enlarged approximately 112,000 times.

second class is made up of a more polar noncollagenous fraction, lacking disaccharides but with a high content of heteropolysaccharide units, including sialic acid, fucose, galactose, mannose, and hexosamine residues. The overall content of the GBM for sialic acid is about 2% (wt/wt).

By virtue of the carbohydrate and anionic moieties, the GBM has been considered to have the nature of a highly hydrated gel, consisting in part of hydrated, polyanionic glycoprotein polymers. The hydrated interstices between the cross-linked mesh of the polymer chains would constitute the pathways for transport of water and solute (5). The permeability characteristics would depend on appropriate spacing of the macromolecular components, i.e. the restrictive dimensions are those of the interstices between the gel fibers or polysaccharide polymers (5).

The Epithelial Cell

The glomerular epithelial cells, or podocytes, are large cells that extend trabecular processes from their cell bodies. These so-called foot processes reach to, and are apparently partially embedded in, the external layer of the lamina rara externa to a depth of about 40–50 nm. The foot processes from adjacent cell bodies interdigitate, the functional significance of which is not known, but it may be involved in maintaining and regulating the caliber of the spaces between the foot processes (the epithelial slits or slit pores) through an integrated contractile mechanism. The foot processes contain large bundles of contractile elements (microfilaments) as well as micro-

tubules, both of which could be involved in maintaining the architecture and the dynamic geometry of the system. The epithelial cell is capable of pinocytosing macromolecules presented to it and of forming digestive or secondary lysosomes.

In between the interdigitating foot processes are the epithelial slits (slit pores or filtration slits), which run from the external aspect of the GBM, and open into the urinary space. They have a width of 20–30 nm at their narrowest point, where they are bridged by a thin diaphragm, 4–6 nm thick in cross section, known as the slit or filtration diaphragm.

Rodewald & Karnovsky (6) showed that this diaphragm, in *en face* view, has a substructure of rod-like units connected in a staggered fashion to a central bar and laterally to the outer leaflets of the plasma membranes of the foot processes. The central bar, the rod-like extensions, and the plasma membranes of the foot processes thus outline rectangular spaces 4×14 nm, which is about the size of an albumin molecule. Consequently, it was suggested (6) that these “pores” may constitute a filtration barrier, and at least for lower forms (7, 8) this might be the case. The chemical composition of the slit diaphragm is unknown, but, as shown by staining methods, it appears to carry considerable negative charge. Farquhar (9) and Farquhar and associates (10), however, consider the slit diaphragms to be special attachment devices, such as long continuous desmosomes of minimal depth, that keep the foot processes in position and maintain the continuity of the epithelial slits as a filtration pathway.

The Distribution of Negative Charge

The distribution of negatively charged groups in the glomerulus has been revealed by the application of polycationic, electron-opaque stains, such as colloidal iron (11), ruthenium red (12), alcian blue (13), colloidal thorium (14), polycationic ferritin (15), and lysozyme (16). Decreased staining after neuraminidase treatment indicates that most of the negative charge as revealed by these stains at low pH is due to sialoprotein. We emphasize the asymmetric, nonuniform distribution of negative charge in the glomerular barrier (Figure 2). This charge distribution is made up of two components, the surface cell coats of the endothelial and epithelial cells and the charge in the GBM proper. The glomerular endothelial cells bear a negatively charged cell coat, which is about 12 nm thick in preparations stained with polycations. The cell coat completely covers the plasma membranes of the cells, including those portions lining the fenestrae.

Both *in situ* (11–14, 17–20) and in isolated GBM (21) there is little binding of cationic stains in the lamina densa but considerable staining in the laminae rarae. This suggests that polyanionic groups are more concentrated in these latter sites, and that the lamina densa is more neutral in

nature, i.e. that it is here that the relatively more nonpolar collagen-like moieties are concentrated, whereas the more polar, noncollagen moieties are concentrated in the laminae rarae.

The epithelial cells carry a notably thick (15–80 nm) polyanionic cell coat on all their surfaces except that of the foot processes directly abutting on the GBM. This coat lines the epithelial slits and extends across, or is a component of, the slit diaphragm.

Mesangial Cells

Mesangial cells, situated in the central axial, centrilobular, and intercapillary regions of the glomerulus, are not prime candidates for the site of a permeability barrier, as they do not form a continuous sheet or structure abutting all the capillaries, which probably would be a necessity for such a role. However, they may play an important secondary role in that: (a) they are thought, by reason of their capacity for spreading, pinocytosis, and phagocytosis (22, 23), to be capable of "cleansing" the GBM and thus maintaining its permeability characteristics; (b) they apparently can take up macromolecules (22–24) that might enter the matrix of the mesangium; and (c) the mesangial cells are thought to have contractile properties and may thus regulate the distribution of blood flow within the glomerulus (23, 25), with possible important consequences for filtration (26–29). The mesangial cells are embedded in a matrix of GBM-like material.

DETERMINANTS OF GLOMERULAR FILTRATION

Physiological Considerations

The permeability properties of the glomerulus and the degree of restriction to passage for different macromolecules is conveniently expressed as the fractional clearance, i.e. the ratio of the clearance of the macromolecule relative to the clearance of inulin. The fractional clearance may vary from zero when the system is impermeable to the macromolecule, to unity when there is no restriction, as for inulin.

SIZE-SELECTIVE FILTRATION It has been known for many years that the glomerulus is a size-selective filter [reviewed in Renkin & Gilmore (28), and Brenner et al (20, 31)]. The fractional clearance decreases with an increase in the effective hydrodynamic molecular radius of the molecule (the Einstein-Stokes radius, or a_e , calculated from the diffusion coefficients for each molecule). At a fractional clearance approaching zero, the radius of the test molecule may be taken as the approximate "pore radius" of the filtration barrier. The estimates of pore size vary from the size of albumin ($a_e = 3.55$ nm) to somewhat larger (4–6 nm); the former is based on the

clearances of globular proteins, the latter on dextrans and polyvinylpyrrolidone (PVP). The explanation for the discrepancies lies in part on the role of molecular charge as a determinant of permeability, and in part on differences in the molecular configuration of the two classes of test molecules. The fractional clearances of neutral dextrans and PVP are severalfold greater than those of globular proteins of comparable molecular size (28). When the clearances of neutral and negatively charged proteins and linear, flexible carbohydrate molecules, matched for charge and molecular size, are compared, the proteins have a lower clearance than the paired carbohydrate probe (32–36). This may be because dextrans and PVP in solutions form loose, hydrated coils, whereas under deforming forces during filtration the flexible molecules elongate, which facilitates their passage through gels (37–42). A similar mechanism has been postulated to occur in the glomerular barrier (28, 43). Compact, cross-linked globular proteins would not undergo these changes during filtration. These differences in clearance behavior, based on molecular configurations, should be taken into account when evaluating and comparing physiological and morphological tracer experiments.

MOLECULAR CHARGE Net molecular charge on macromolecules is a prime determinant of their glomerular clearances, which neatly relates to the high fixed negative charge (FNC) in the glomerular barrier (see above). As shown by Brenner and associates (32, 34, 44) the fractional clearances of negatively charged dextran sulfates are less than those for neutral dextrans of equivalent molecular sizes. This applies over a wide range, up to $a_e = 42$ nm or more. For molecules the size of albumin ($a_e = 3.6$ nm), dextran sulfate has a lower fractional clearance than neutral dextran: this implies that the anionic nature of albumin ($pI = 4.7$) as well as its size is an important determinant of permeability. Conversely, positively charged diethylaminoethyl dextrans have higher clearances than the equivalently sized neutral or anionic dextrans (45). Similar observations were made for the globular protein horseradish peroxidase: cationic, neutral, and anionic HRP, all of similar size, showed a progressive decrease in fractional clearances for each category (36).

In some glomerulopathies, there is a decrease in FNC in the various components of the glomerular barrier (19, 46–53). The mechanism of the proteinuria has been well studied in the experimental models nephrotoxic serum nephritis (NTN) (46, 54) and puromycin aminonucleoside nephrosis (PAN) (55) in which there is loss of FNC (19, 46, 48–50, 52). There was increased fractional clearance for dextran sulfates relative to neutral dextrans, which suggests a loss of charge discrimination in the barrier, now only size selective and not highly restrictive for molecules the size of albu-

min. In NTN Kreisberg & Karnovsky (53) reported a loss of FNC before the onset of proteinuria, which supports the concept that such loss leads to loss of charge selectivity.

Seiler et al (56, 57) perfused polycations through the glomerulus, and found spreading of foot processes, obliteration of slit pores, and reduction of FNC—common changes in pathological states with proteinuria. Polyanions and neutral molecules had no such effect, and the changes induced by polycations were reversible with polyanions. It can be hypothesized that FNC not only regulates permeability for polyanions (32, 34, 44) but also maintains, by electrostatic repulsive forces, the structural integrity of the epithelial layer (19, 56, 57). Glomerular polyanion content may also be a determinant of water permeability, through regulation of foot process geometry: Epithelial foot process spreading and obliteration of epithelial slits have been imputed to lead to a decrease in the filtration surface area (58), and in rat PAN, in which these morphological changes are prominent, there is a decrease in GFR and the ultrafiltration coefficient (55). On morphometric grounds, the aggregate area of the pores in the slit diaphragm may be a major determinant of water permeability (6, 28, 59).

HEMODYNAMIC FACTORS The fractional clearance for macromolecules shows an inverse relation to the filtration of water (the glomerular filtration rate or GFR): an increase in GFR gives a decreased fractional clearance, although net macromolecular transport is increased, and vice versa. This is a reflection of so-called molecular sieving (60, 61). Factors that in part determine the GFR are the effective filtration pressure and the glomerular plasma flow. A decrease in one or both of these factors could yield a decreased GFR, with a relative decrease in convection and increase in diffusion, of molecules across the filter. The converse also applies. Glomerular hemodynamics functionally modulates macromolecular transport by complex mechanisms [see reviews of Brenner and associates (30, 35, 44, 62)]. As examples, in volume expansion with isoncotic plasma in rats, a decreased fractional clearance was observed (34), and with intravenous infusion of angiotensin II in rats, associated with a decreased glomerular plasma flow, fractional clearances increased (29).

ULTRASTRUCTURAL TRACER STUDIES

Over the past fifteen years or so, morphologists have attempted to localize the exact ultrastructural location of the primary filtration barrier. The rationale has been to utilize intravenously injected electron-opaque molecules (or molecules that can be rendered electron opaque) and to observe at what level in the glomerular barrier they are restricted from passage into

the urine. In early studies, Farquhar et al (10) found that ferritin ($a_e = 6.1$ nm) was essentially retarded at the level of the GBM, and that this was the main barrier to macromolecular filtration. Graham & Karnovsky (63) subsequently found that native horseradish peroxidase ($a_e = 3.0$ nm) readily traversed the GBM, with only slight apparent retardation, and passed through the epithelial slits into the urinary space. From physiological data, little retardation would be expected for a molecule of this size. This study established that the route whereby lipid-insoluble macromolecules reach the urinary space is essentially extracellular. However, these authors also found that human myeloperoxidase ($a_e = 6.0$ nm) rapidly crossed the GBM, but was apparently restricted from passing into the urine (restriction would be expected for a molecule of this size) at the level of the slit diaphragm. From these observations it was concluded that the final barrier for molecules such as albumin and globulin was at the level of the epithelial slit, probably at the slit diaphragm, whereas extremely large molecules such as ferritin were retarded by the GBM. In a later study Venkatachalam et al (64) utilized catalase ($a_e = 5.2$ nm) as a probe, and found that it slowly crossed the GBM, displaying a concentration gradient and finally reaching the level of the slit diaphragm where it was apparently restricted from entering the urinary space. In the light of these and the previous observations, the "double barrier" hypothesis was proposed: the GBM acts as a coarse filter, the slit diaphragm as a finer one (63-66).

This seemed a neat hypothesis, since it accounted for the various and discrepant ultrastructural observations and also explained some aspects of selectivity in proteinuria, but it should be realized that the hypothesis was put forward at a time when the glomerular barrier was considered as a porous size-selective filter. With our current knowledge of the role of charge, molecular shape, and the effects of hemodynamic changes, the double barrier hypothesis based solely on size discrimination is obsolete. Indeed, Venkatachalam et al (64), noting the differences in behavior of the relatively large molecules myeloperoxidase and catalase postulated that these differences might also be attributable to variations in molecular charge (for catalase the $pI = 5.7$, for myeloperoxidase the $pI > 10.0$).

Recently, Caulfield & Farquhar (9, 67), utilizing neutral dextrans as probes, found no evidence of restriction at the slit pore level, but rather at the level of the GBM; this reinforced the original view of Farquhar et al (10) that the GBM is the primary barrier.

In regard to charge discrimination at the ultrastructural tracer level, Rennke, Cotran & Venkatachalam (68) and Rennke & Venkatachalam (69) strikingly showed that ferritin molecules of essentially the same size, but bearing varying charge, penetrate differentially into the glomerular barrier. Native anionic ferritin ($pI = 4.1-4.7$) was almost completely excluded from

entering the GBM, whereas with progressive cationization the molecules penetrated further across the GBM, the most cationic penetrating up to the level of the slit pore. The prior and discrepant results obtained with anionic ferritin and catalase on the one hand and cationic myeloperoxidase on the other are not explicable in terms of their differences in net charge (31, 44, 70, 71).

Investigating the level of restriction for endogenous macromolecules, such as albumin, Ryan & Karnovsky (58) applied drip fixation *in situ* to the superficial glomeruli of Munich-Wistar rats. The glomeruli were fixed as far as possible under circumstances approximating normal hemodynamics. They found that endogenous plasma albumin was restricted at the endothelial and subendothelial level. As albumin is an anionic molecule ($pI = 4.7$), this result is well in accord with those related to native anionic ferritin ($pI = 4.1-4.7$). Similar results were obtained for immunoglobulin ($pI = 6.3-7.3$) and exogenous catalase ($pI = 5.7$) (72). However, when the glomeruli were fixed under low flow conditions, very different results were obtained. Albumin passed rapidly across the GBM and the slit diaphragm, entering the urinary space (58), whereas catalase and immunoglobulin (72) traversed the GBM up to the level of the slit diaphragm but did not enter the urinary space. These patterns were rapidly reversed on restoring normal flow. Furthermore, the same low flow patterns of passage of the macromolecules into the urinary space or up to the slit diaphragm were obtained with immersion fixation. Thus it appears that, with reduction in blood flow, diffusion of the tracers occurs across the glomerular barrier, which is in concordance with physiological data (see above), and that the slit diaphragm is a barrier for molecules larger than albumin, if they reach that level. The prior results of anionic catalase slowly, and cationic myeloperoxidase rapidly, reaching the slit diaphragm are thus also explicable not only on the basis of charge differences but also on the basis of the fact that the original experiments were run under no-flow conditions, *i.e.* with immersion fixation.

BARRIER FUNCTION AND STRUCTURE

In retrospect it now appears unreasonable to attempt to assign barrier function solely to any one component of the glomerular capillary wall. For polyanions, such as albumin, the endothelial cell coat and the charged subendothelial components would seem to be the major barrier, with a possibly important contribution from the heavily charged cell coat of the epithelial cells and the dense charge in the lamina rara externa operating through electrostatic field effects. For cationic macromolecules the slit diaphragm would be the final barrier for molecules larger than about 3.6 nm.

With decrease in GFR diffusing macromolecules would be retarded at the same level, i.e. in the proximity of the slit diaphragm. Neutral molecules would be sieved by the GBM according to size. Finally, the epithelial and mesangial cells, by virtue of their capacity to take up macromolecules by pinocytosis or phagocytosis, would play a secondary monitoring role.

Literature Cited

1. Latta, H. 1973. Ultrastructure of the glomerulus and the juxtaglomerular apparatus. In *Handbook of Physiology, Sect. 8, Renal Physiology*, ed. J. Orloff, R. W. Berliner, pp. 1-29. Washington, DC: Am. Physiol. Soc. 1082 pp.
2. Kefalides, N. A. 1972. Biochemical studies of the glomerular basement membrane in the normal kidney. *Adv. Nephrol. Necker Hosp.* 2:3-24
3. Kefalides, N. A. 1973. Structure and biosynthesis of basement membranes. *Int. Rev. Connect. Tissue Res.* 6:63-104
4. Spiro, R. G. 1972. Basement membranes and collagens. In *Glycoproteins: Their Composition, Structure and Function*, ed. A. Gottschalk, pp. 964-99, Pt. B. Amsterdam: Elsevier. 731 pp.
5. Laurent, T. C. 1970. The structure and function of the intercellular polysaccharides in connective tissue. In *Alfred Benzon Symposium II. Capillary Permeability: The Transfer of Molecules and Ions Between Capillary Blood and Tissue*, ed. C. Crone, N. A. Lassen, pp. 261-77. Copenhagen: Munksgaard. 681 pp.
6. Rodewald, R., Karnovsky, M. J. 1974. Porous substructure of the glomerular slit diaphragm in the rat and mouse. *J. Cell Biol.* 28:423-33
7. Schaffner, A., Rodewald, R. 1978. Glomerular filtration of ferritin in the bullfrog, *Rana catesbeiana*. *J. Cell Biol.* In press
8. Schaffner, A., Rodewald, R. 1978. Filtration barriers in the coelomic sac of the crayfish, *Procambarus clarkii*. *J. Ultrastruct. Res.* In press
9. Farquhar, M. G. 1975. The primary glomerular filtration barrier—basement membrane or epithelial slits? Editorial Review. *Kidney Int.* 8:197-211
10. Farquhar, M. G., Wissig, S. L., Palade, G. E. 1961. Glomerular permeability: I. Ferritin transfer across the normal glomerular capillary wall. *J. Exp. Med.* 113:47-66
11. Jones, D. B. 1969. Mucosubstances of the glomerulus. *Lab. Invest.* 21:119-25
12. Latta, H., Johnston, W. H., Stanley, T. M. 1975. Sialoglycoproteins and filtration barriers in the glomerular capillary wall. *J. Ultrastruct. Res.* 51:354-76
13. Behnke, O., Zelandier, T. 1970. Preservation of intercellular substances by the cationic dye alcian blue in preparative procedures for electron microscopy. *J. Ultrastruct. Res.* 31:424-38
14. Rambourg, A., Leblond, C. P. 1967. Electron microscope observations on the carbohydrate rich cell coat present at the surface of cells in the rat. *J. Cell Biol.* 32:27-53
15. Seiler, M. W., Rennke, H. G., Cotran, R. S., Venkatachalam, M. A. 1976. Distribution of glomerular anionic sites as revealed by polycationic compounds. *Am. J. Pathol.* 82:54a (Abstr.)
16. Caulfield, J. P., Farquhar, M. G. 1976. Distribution of anionic sites in glomerular basement membranes: Their possible role in filtration and attachment. *Proc. Natl. Acad. Sci. USA* 73:1646-50
17. Groniowski, J., Biczyskova, W., Wal-ski, M. 1969. Electron microscope studies on the surface coat of the nephron. *J. Cell Biol.* 40:585-601
18. Latta, H., Johnston, W. H. 1976. The glycoprotein inner layer of glomerular capillary basement membrane as a filtration barrier. *J. Ultrastruct. Res.* 57:65-67
19. Michael, A. F., Blau, E., Vernier, R. L. 1970. Glomerular polyanion alteration in aminonucleoside nephrosis. *Lab. Invest.* 23:649-57
20. Mohos, S. C., Skoza, L. 1969. Glomerular sialoprotein. *Science* 164:1519-21
21. Nicholes, B. K., Krakower, C. A., Greenspon, S. A. 1973. The chemically isolated lamina densa of the renal glomerulus. *Proc. Soc. Exp. Biol. Med.* 142:1316-21
22. Farquhar, M. G., Palade, G. E. 1962. Functional evidence for the existence of a third cell type in the renal glomerulus. Phagocytosis of filtration residues by a distinctive "third" cell. *J. Cell Biol.* 13:55-87
23. Latta, H., Maunsbach, A. B., Cook, M. L. 1962. Relations of the centrolobular region of the glomerulus to the juxta-

- glomerular apparatus. *J. Ultrastruct. Res.* 6:562-78
24. Graham, R. C., Karnovsky, M. J. 1966. The early stage of absorption of injected horseradish peroxidase in the proximal tubule of the mouse kidney. Ultrastructural cytochemistry by a new technique. *J. Histochem. Cytochem.* 14:291-302
 25. Zamboni, L., de Martino, C. 1968. A reevaluation of the mesangial cells of the renal glomerulus. *Z. Zellforsch. Mikrosk. Anat.* 86:364-83
 26. Barger, A. C., Herd, J. A. 1971. The renal circulation. *N. Engl. J. Med.* 284:482-90
 27. Barger, A. C., Herd, J. A. 1973. Renal vascular anatomy and distribution of blood flow. See Ref. 1, pp. 249-313
 28. Renkin, E. M., Gilmore, J. P. 1973. Glomerular filtration. See Ref. 1, pp. 185-248
 29. Bohrer, M. P., Deen, W. M., Robertson, C. R., Brenner, B. M. 1977. Mechanism of angiotensin II-induced proteinuria in the rat. *Am. J. Physiol.* 2:F13-F21
 30. Brenner, B. M., Deen, W. M., Robertson, C. R. 1976. Glomerular filtration. In *The Kidney*, Vol. 1, ed. B. M. Brenner, F. C. Rector, pp. 251-71. Philadelphia: Saunders. 797 pp.
 31. Brenner, B. M., Hostetter, T. H., Humes, H. D. 1978. Molecular basis of proteinuria of glomerular origin. *N. Engl. J. Med.* 298:826-33
 32. Chang, R. L. S., Deen, W. M., Robertson, C. R., Brenner, B. M. 1975. Permeability of the glomerular capillary wall: III. Restricted transport of polyanions. *Kidney Int.* 8:212-18
 33. Chang, R. L. S., Robertson, C. R., Deen, W. M., Brenner, B. M. 1975. Permeability of the glomerular capillary wall to macromolecules: I. Theoretical considerations. *Biophys. J.* 15:861-86
 34. Chang, R. L. S., Ueki, I. F., Troy, J. L., Deen, W. M., Robertson, C. R., Brenner, B. M. 1975. Permeability of the glomerular capillary wall to macromolecules: II. Experimental studies in rats using neutral dextran. *Biophys. J.* 15:887-906
 35. Deen, W. M., Bohrer, M. P., Robertson, C. R., Brenner, B. M. 1977. Determinants of the transglomerular passage of macromolecules. *Fed. Proc.* 36:2614-18
 36. Rennke, H. G., Patel, Y., Venkatachalam, M. A. 1978. Glomerular filtration of proteins: Clearance of anionic, neutral, and cationic horseradish peroxidase in the rat. *Kidney Int.* 13:324-28
 37. Cerf, R., Scheraga, H. A. 1952. Flow birefringence in solutions of macromolecules. *Chem. Rev.* 51:185-261
 38. Ogston, A. G., Woods, E. F. 1953. Molecular configuration of dextrans in aqueous solution. *Nature* 171:221-22
 39. Laurent, T. C., Pietruszkiewicz, A. 1961. The effect of hyaluronic acid on the sedimentation rate of other substances. *Biochim. Biophys. Acta* 49:258-64
 40. Laurent, T. C., Bjork, I., Pietruszkiewicz, A., Persson, H. 1963. On the interaction between polysaccharides and other macromolecules. II. The transport of globular particles through hyaluronic acid solutions. *Biochim. Biophys. Acta* 78:351-59
 41. deGennes, P. G. 1971. Reptation of a polymer chain in the presence of fixed obstacles. *J. Chem. Phys.* 55:572-79
 42. Laurent, T. C., Preston, B. N., Pertoft, H., Guftafsson, B., McCabe, M. 1975. Diffusion of linear polymers in hyaluronate solutions. *Eur. J. Biochem.* 53:129-36
 43. Renkin, E. M. 1970. Permeability and molecular size in "peripheral" and glomerular capillaries. See Ref. 5, pp. 544-47
 44. Brenner, B. M., Bohrer, M. P., Baylis, C., Deen, W. M. 1977. Determinants of glomerular permselectivity: Insights derived from observations in vivo. Editorial Review. *Kidney Int.* 12:229-37
 45. Bohrer, M. P., Humes, H. D., Baylis, C., Robertson, C. R., Brenner, B. M. 1977. Facilitated transglomerular passage of circulating polycations. *Clin. Res.* 25:505A
 46. Bennett, C. M., Glasscock, R. J., Chang, R. L. S., Deen, W. M., Robertson, C. R., Brenner, B. M. 1976. Permeability of the glomerular capillary wall: Studies of experimental glomerulonephritis in the rat using dextran sulfate. *J. Clin. Invest.* 57:1287-94
 47. Blau, E. B., Hass, J. E. 1973. Glomerular sialic acid and proteinuria in human renal disease. *Lab. Invest.* 28:477-81
 48. Blau, E., Michael, A. F. 1971. Rat glomerular basement membrane composition and metabolism in aminonucleoside nephrosis. *J. Lab. Clin. Med.* 77:97-109
 49. Chiu, J., Drummond, K. N. 1972. Chemical and histochemical studies of glomerular sialoprotein in nephrotoxic nephritis in rats. *Am. J. Pathol.* 68:391-406

50. Chiu, J. M., Smith, M., Drummond, K. N. 1973. Proteinuria and glomerular sialoprotein (GSP) in experimental renal disease. *Clin. Res.* 21:681 (Abstr.)
51. Couser, W. G., Stilmant, M. M., Darby, C. 1976. Autologous immune complex nephropathy: I. Sequential study of immune complex deposition, ultrastructural changes, proteinuria and alterations in glomerular sialoprotein. *Lab. Invest.* 34:23-30
52. Kalant, H., Misra, R. P., Manley, R. St. J., Wilson, J. 1966. Glomerular basement membrane in experimental nephrosis: x-ray diffraction and electrophoretic studies. *Nephron* 3:167-72
53. Kreisberg, J. I., Karnovsky, M. J. 1977. Rapid and focal loss of negative charge (NC) following administration of nephrotoxic serum in rats. *Kidney Int.* 12:515 (Abstr.)
54. Chang, R. L. S., Deen, W. M., Robertson, C. R., Bennett, C. M., Glasscock, R. J., Brenner, B. M. 1976. Permeability of the glomerular capillary wall: Studies of experimental glomerulonephritis in the rat using neutral dextran. *J. Clin. Invest.* 57:1272-86
55. Baylis, C., Bohrer, M. P., Troy, J. L., Robertson, C. R., Brenner, B. M. 1976. Mechanisms of the puromycin induced defects in transglomerular passage of water and macromolecules. *Kidney Int.* 10:554 (Abstr.)
56. Seiler, M. W., Venkatachalam, M. A., Cotran, R. S. 1975. Glomerular epithelium: Structural alterations induced by polycations. *Science* 189:390-93
57. Seiler, M. W., Rennke, H. G., Venkatachalam, M. A., Cotran, R. S. 1977. Pathogenesis of polycation induced alterations ("fusion") of glomerular epithelium. *Lab. Invest.* 36:48-61
58. Ryan, G. B., Karnovsky, M. J. 1976. Distribution of endogenous albumin in the rat glomerulus. Role of hemodynamic factors in glomerular barrier function. *Kidney Int.* 9:36-45
59. Shea, S. M., Morrison, A. B. 1975. A stereological study of the glomerular filter in the rat. Morphometry of the slit diaphragm and basement membrane. *J. Cell Biol.* 67:436-43
60. Pappenheimer, J. R. 1953. Passage of molecules through capillary walls. *Physiol. Rev.* 33:387-423
61. Pappenheimer, J. R. 1955. Über die permeabilität der glomerulummembranen in der niere. *Klin. Wochenschr.* 33:362-65
62. Brenner, B. M., Baylis, C., Deen, W. M. 1976. Transport of molecules across renal glomerular capillaries. *Physiol. Rev.* 56:502-34
63. Graham, R. C., Karnovsky, M. J. 1966. Glomerular permeability. Ultrastructural cytochemical studies using peroxidases as protein tracers. *J. Exp. Med.* 124:1123-34
64. Venkatachalam, M. A., Karnovsky, M. J., Fahimi, H. D., Cotran, R. S. 1970. An ultrastructural study of glomerular permeability using catalase and peroxidase as tracer proteins. *J. Exp. Med.* 132:1153-67
65. Karnovsky, M. J., Ainsworth, S. K. 1972. The structural basis of glomerular filtration. *Adv. Nephrol.* 2:35-60
66. Karnovsky, M. J., Venkatachalam, M. A., Graham, R. C. Jr., Cotran, R. S. 1969. Ultrastructural basis of glomerular permeability to macromolecules. *Proc. IV Int. Congr. Nephrol., Stockholm, 1969.* 1:81-82
67. Caulfield, J. P., Farquhar, M. G. 1974. The permeability of glomerular capillaries to graded dextrans. *J. Cell Biol.* 63:883-903
68. Rennke, H. G., Cotran, R. S., Venkatachalam, M. A. 1975. Role of molecular charge in glomerular permeability. Tracer studies with cationized ferritins. *J. Cell Biol.* 67:638-46
69. Rennke, H. G., Venkatachalam, M. A. 1977. Glomerular permeability: In vivo tracer studies with polyanionic and polycationic ferritins. *Kidney Int.* 11:44-53
70. Venkatachalam, M. A., Rennke, H. G. 1978. The structural and molecular basis of glomerular filtration. *Circ. Res.* 43:337-47
71. Karnovsky, M. J. 1979. The structural bases for glomerular filtration. In *International Academy of Pathology, Monograph 20: Renal Disease: Present Status*, ed. J. Chung, B. Spargo, F. Mostofi, N. Abell. Baltimore: Williams & Wilkins. In press
72. Ryan, G. B., Hein, S. J., Karnovsky, M. J. 1976. Glomerular permeability to proteins. Effect of hemodynamic factors on the distribution of endogenous immunoglobulin and exogenous catalase in the rat glomerulus. *Lab. Invest.* 34:415-27

LOCALIZATION AND ANTIBIOTIC MANAGEMENT OF URINARY TRACT INFECTION

◆7318

*Leslie S. T. Fang, Ph.D., M.D., Nina E. Tolckoff-Rubin,
M.D., and Robert H. Rubin, M.D.*

Department of Medicine, Harvard Medical School, and the Infectious Disease
and Renal Units, Medical Service, Massachusetts General Hospital,
Boston, Massachusetts 02114

Over the past three decades the major question dominating the field of urinary tract infection has been, "What is the contribution of such infections to the development of chronic renal failure?" The question was initially posed because of the frequent demonstration at autopsy or at the time of transplantation of the pathologic changes of chronic pyelonephritis—even in patients without a clearcut history of documented urinary tract infection. Because of these pathologic findings, the hypothesis arose that symptomatic or asymptomatic bacteriuria could cause progressive parenchymal damage and loss of renal function. However, in recent years at least three different lines of evidence have developed, demonstrating that in adults infection is essentially never the primary cause of chronic renal failure (1):

1. The pathologic changes of chronic pyelonephritis are caused by a number of different disease processes, ranging from chronic analgesic abuse and other intoxications to long-standing ischemia (Table 1). Therefore, the term now used to describe these pathologic findings is *interstitial nephritis*, and only in those cases of interstitial nephritis in which clearcut evidence of infection is present is the diagnosis of chronic pyelonephritis made (1, 2).

2. Long-term clinical studies of both adult males (3) and females (1, 4, 5) with chronic or recurrent bacteriuria have not demonstrated progressive renal damage in the absence of significant anatomical defects.

Table 1 Etiologies of chronic interstitial nephritis

Analgesic abuse	Nephrolithiasis	Radiation
Hypertension	Hypercalcemia	Lead nephropathy
Renal vascular disease	Hypokalemia	Balkan nephropathy
Obstruction	Sickle cell anemia	

3. A retrospective study (6) of 100 consecutive cases of renal failure caused by chronic interstitial nephritis seen at a major medical center failed to reveal a single case in which infection was the primary cause of the renal disease. However, in the presence of the conditions outlined in Table 1, particularly vesicoureteral reflux and obstruction, infection was thought to play an important secondary role. Given this body of evidence, then, the major concerns in the study of adult urinary tract infection today are its pathogenesis and clinical management.

Among adult women, urinary tract infection is the most common of all bacterial infections, affecting 10–20% of women at some time in their lives (7), with an estimated 6% of the adult female population developing urinary tract infections each year (8). Treatment of these infections with a wide variety of antimicrobial agents results in bacteriologic control in approximately 85% of such infections, but with an unacceptably high rate of recurrence (9). This review, then, is devoted to a consideration of those factors that will lead to the more effective clinical management of this rather universal affliction of the adult female population.

PATHOGENESIS OF URINARY TRACT INFECTION IN WOMEN

Stamey (10) and his colleagues showed that colonization of the vaginal introitus by members of the *Enterobacteriaceae*, particularly *Escherichia coli*, derived from the fecal reservoir is the essential first step in the production of bacteriuria. Women with recurrent urinary tract infections have a statistically higher rate of introital carriage with *Enterobacteriaceae* than do women resistant to urinary tract infections (11). Vaginal mucosal cells of susceptible women appear to have a greater affinity for these bacteria than do such cells from resistant women (12). This difference in adherence may, at least in part, be related to an absence of antibodies directed against these organisms in the cervicovaginal secretions of these susceptible women (13).

Entry of bacteria into the bladder through the relatively short female urethra can occur spontaneously, but urethral trauma has long been incriminated as playing a major role in such entry. Bran et al (14), performing suprapubic aspiration in anesthetized patients undergoing gynecological procedures, found a significant increase in the recovery of bacteria from

these aspirates after urethral milking. The organisms retrieved in the bladder aspirates correlated quite well with those simultaneously demonstrated to be present in the distal urethra. Similarly, serial determinations of urine bacterial counts before and after sexual intercourse also demonstrated significant increases in bacterial counts following 30% of the intercourse episodes. Again, the positive urine cultures reflected the prevailing distal urethral flora, and it was concluded that transient bacteriuria occurs frequently in the sexually active female (15). The assumption has been made that, in the urinary-tract-infection-resistant woman, the organisms introduced into the bladder with such urethral manipulation are the diphtheroids, *Staphylococcus epidermidis*, lactobacilli, etc that normally colonize the distal urethra—rare causes of urinary tract infection. The approximately 20% of women who are susceptible to the development of true urinary tract infection have *E. coli* and/or other gram negative bacilli colonizing the area, so that a significant inoculum of these more virulent organisms may be introduced with intercourse, and clinical infection develops.

Following the entry of potential pathogens into the urine, a number of host defense factors act together to decrease the likelihood of more than transient infection (16). Normal voiding will eliminate some organisms, while the bladder mucosal surface itself, through mucus trapping of the organisms and polymorphonuclear leukocyte response, has antibacterial properties. In addition, urine of low pH, high or very low osmolarity, high urea concentration, and high organic acid content, supports bacterial growth quite poorly. Even when infection is sustained, the site of involvement is frequently restricted to the bladder, but in approximately 30–50% of instances there is further extension of the infection into and up the ureter to the kidney (10). Although urine usually does not reflux into the ureters from an uninfected bladder, cystitis itself, in addition to anatomic defects, may produce reflux. Once bacteria are introduced into the ureter, they may ascend to the kidney unaided, but this would be greatly increased by any process that interferes with normal ureteral peristaltic function. Gram negative bacteria and their endotoxins have a marked antiperistaltic effect, as do pregnancy and ureteral obstruction (1, 10).

Once infected urine gains access to the renal pelvis, it can enter the renal parenchyma via the ducts of Bellini at the papillary tips, and then spread outward along the collecting tubules. Clearly, this whole process is hastened and exacerbated by any element of significant vesicoureteral reflux. This is particularly so since even moderate degrees of such reflux may be associated with intrarenal reflux, thus escalating the process further (1, 2).

From a therapeutic point of view, this pathogenetic scheme suggests an important difference between infection confined to the bladder and that

extending into the kidneys. In particular, it needs to be emphasized that infection within the bladder primarily involves the superficial mucosa, and is constantly bathed by urine in which exceedingly high concentrations of effective antimicrobial agents may be easily achieved. Not surprisingly, there is little immunologic response to such superficial infection. In contrast, renal infection is a deep tissue infection in which a significant immunologic response is the rule, and where delivery of effective concentrations of antimicrobial agents may be more of a problem. Therefore, the definition of the anatomic site of urinary tract infection is of more than academic interest.

LOCALIZATION OF THE SITE OF URINARY TRACT INFECTION

Clinical Evaluation

The ability of the clinician to distinguish between kidney infection and infection restricted to the bladder on the basis of presenting signs and symptoms is quite limited. Not only will many patients with renal infection present only with complaints referable to the lower urinary tract (dysuria, frequency, and urgency), but also some patients with bladder infection will complain of flank pain, fever, and systemic symptoms usually associated with pyelonephritis (Table 2). In addition, patients with the so-called acute urethral syndrome (symptomatic abacteriuria) will have similar symptoms in the face of urine cultures showing insignificant or no growth of bacteria

Table 2 Acute urinary tract infection in general practice (in %)^a

Manifestations	No bacteriuria or <10,000 organisms per ml (23 patients)	Renal bacteriuria (21 patients)	Bladder bacteriuria (22 patients)
Symptoms suggesting lower urinary tract infection:			
Frequency	95	98	70
Burning	70	68	70
Suprapubic pain	70	68	51
Symptoms suggesting upper urinary tract infection:			
Flank pain	50	48	19
Fever	35	44	4
Rigors	15	32	15
Nausea and vomiting	25	24	8
Hematuria (macroscopic)	25	20	12

^aFrom (8).

(8, 10, 17-19). Better techniques of infection localization are clearly required to fairly evaluate the hypothesis that site of infection affects the clinical course and the response to therapy (8).

Invasive Techniques

The only direct method to localize infection site is bilateral ureteral catheterization (10, 19). The technique involves cystoscopy, is at times technically difficult, and is too invasive for general application. Nonetheless, it provides a standard against which all other methods of localization must be compared. In an effort to employ a less invasive, more generally applicable procedure, Fairley et al (18, 20) in the mid-1960s introduced a bladder washout procedure to localize the site of infection. In this procedure, a Foley catheter is introduced into the bladder; the bladder is irrigated with an antibiotic solution (usually neomycin or neomycin-polymyxin); serial samples are then collected. Patients with lower tract infection will have sterile urine during the collection period following the washout, while patients with renal infection will have bacteria in all of the post-washout samples.

This method compares favorably with ureteral catheterization, does not require expert urologic aid, and has less potential morbidity. There may be some inaccuracies with this method, however. Some authors (21) suggest that misleading results may be obtained due to transient vesicoureteral reflux in severe cystitis, with false-positive results being obtained in such patients. Others (22) point out that a constant outflow of bacteria into the renal pelvis is required if counting of bacteria in ureteral urine is to be used as an indication of renal infection. This, however, is not always the case and intermittent discharge of infected urine from renal parenchyma may give false negative results. In addition, unilateral renal involvement cannot be distinguished from bilateral involvement. Still, in practical terms the reliability of this method is quite satisfactory, and most noninvasive techniques of infection localization are now compared to this technique, rather than to ureteral catheterization.

Noninvasive Techniques

Three types of noninvasive techniques have been employed in an attempt to differentiate between renal and bladder infection: assay of renal medullary function by measurement of maximal urinary concentrating capacity; measurement of urinary enzymes as an index of tissue invasion and inflammation; and the determination of the immunologic response to the infection. Implicit in the rationale behind each of these tests has been the recognition of the pathophysiologic differences between renal and bladder infection: the former, deep medullary infection at an anatomic site where

critical aspects of urine formation are taking place, and where inflammatory and immunologic response is brisk and extensive; the latter, superficial mucosal, at an anatomic site where little is occurring functionally, and where both inflammatory and immunologic response is limited.

URINARY CONCENTRATING ABILITY Patients with upper tract infection differ from those with lower tract infections in that they often have an abnormality in renal concentrating ability. The inability to concentrate is most obvious when maximal urinary concentrating ability is examined (23-27). In one representative study, Ronald et al (27) studied the maximal concentrating ability in 38 patients whose site of infection was directly localized by ureteral catheterization. They demonstrated that renal but not bladder bacteriuria was associated with a decreased concentrating ability. Bilateral infection was associated with a greater defect than unilateral infection. In patients with unilateral infection, the infected kidney fails to concentrate, while the uninfected kidney concentrates normally. With eradication of infection, concentrating ability improves. Even in this study, however, the overlap in maximal urinary osmolality achieved in patients with bladder, unilateral, and bilateral renal infections is considerable. The test is therefore neither a sensitive nor a consistently reliable method of localization.

MEASUREMENT OF URINARY ENZYMES Wacker & Dorfman (28) found that urinary lactic acid dehydrogenase (LDH) activity was elevated in 25% of patients with pyelonephritis. However, in addition to false negative results, the presence of blood and heavy proteinuria will cause false positive results. Urinary beta-glucuronidase activity has been found to be higher in patients with pyelonephritis than in patients with lower tract infection (29). Ronald et al (30), using ureteral catheterization for localization, found slightly higher beta-glucuronidase activity in patients with renal infection than in controls or patients with bladder infection. However, considerable overlap was seen in these three groups, and the test was not useful in the individual patient. Turck (31) found a similar overlap problem. Urinary enzymes have, by and large, been unreliable and insensitive indicators of the anatomy of urinary tract infection.

MEASUREMENT OF C-REACTIVE PROTEIN Jodal and co-workers (32, 33), reported that consistently elevated levels of C-reactive protein in serum, as detected by an immunodiffusion technique, were seen in children with acute pyelonephritis. Children with acute cystitis, on the other hand, did not have elevated C-reactive protein levels. Sequential determination of C-reactive protein values in children with pyelonephritis showed that ap-

appropriate therapy led to a progressive decrease in these levels, while ineffective therapy resulted in a slow decrease or even an increase in the levels. However, localization of infections in these studies was made primarily on clinical grounds, and the assigned diagnosis did not correlate with results of bladder washout studies in 5 of 25 patients studied. The C-reactive protein level may also be elevated in a wide variety of other clinical conditions, and false positive values may be observed (34).

MEASUREMENT OF ANTIBODY RESPONSES TO THE BACTERIA Renal infection is associated with the net synthesis of specific antibody directed against antigens of the infecting organism (35). Cotran (36), using fluorescent antibody techniques, demonstrated gamma-globulin-containing cells in the kidney during the course of experimental pyelonephritis. Lehrman et al (37), injected *E. coli* into rabbits with one ureter transiently occluded and studied the local immune response. Using in vitro incubation techniques, they were able to demonstrate protein synthesis in the pyelonephritic kidney, with most of this protein being IgG class globulin, significant portions of which were specific antibodies directed against the infecting strain of organism.

Given this information, it is not surprising that immunologic techniques have been applied to the problem of infection localization. Percival et al (38), using a bacterial agglutination test, found elevated serum antibody levels in patients with symptoms of acute pyelonephritis, with these titers falling in response to antibiotic therapy. Patients with clinically inapparent pyelonephritis also had high antibody levels, while those with infection limited to the lower tract had normal titers. Clark et al (39), localized the site of infection by ureteral catheterization and examined hemagglutinating antibody response and confirmed that some patients with renal infection had higher hemagglutinating antibody titers than patients with bladder bacteriuria. However, a wide range of titers and a considerable overlap between the two groups were seen, so that the determination of antibody titer was of limited usefulness in the individual patient (31).

More promising, and presently the most accurate of the indirect techniques for localizing the site of infection, is the demonstration of antibody coating of bacteria in the urine. Thomas (40), Jones (41) and their colleagues, using an immunofluorescence assay, showed that bacteria originating from the kidney were coated with antibody, while bacteria associated with lower urinary tract infection were antibody negative. In the Jones study (41), in particular, there was at least 95% correlation with the results of bladder washout tests (Table 3). The high degree of sensitivity and reliability of this assay were confirmed by other workers (9, 42, 43). However, both false positive and false negative results may occur. Vaginal (44)

Table 3 The relationship of the site of infection (determined by direct localization techniques) to the presence of antibody-coated bacteria in the urine sediment^a

Site	Antibody-coated bacteria	
	Present	Absent
Upper tract	17	1
Lower tract	0	8

^aModified from (41).

or rectal flora (45) contaminating a urine specimen may give false positive results. Patients with heavy proteinuria may also have false positive results (46). In males, prostatitis may also be associated with the antibody coating of the infecting bacteria (47). False negative results in adults may occur early in the course of renal infection, during the time necessary to generate the antibody response. In experimental hematogenous pyelonephritis in the rabbit, Smith et al (48), showed a time lag of 11–15 days between introduction of the organism and the development of a positive assay. Finally, for reasons as yet unclear, the test appears to be unreliable in children (49).

With the realization of these limitations, the test can be a reliably sensitive and noninvasive method of localizing the site of infection in adults. This is the only reasonably accurate noninvasive assay system currently available for localizing the anatomic site of urinary tract infection.

CORRELATION BETWEEN THE SITE OF INFECTION AND THE RESPONSE TO ANTIMICROBIAL AGENTS

Pathophysiologically, it is clear that there is a major difference between renal and bladder urinary tract infections. Clinically, evidence that such distinctions are of practical importance comes from two sources: an analysis of the pattern of recurrence in those patients with recurrent bacteriuria; and an analysis of the response to antimicrobial agents in patients whose infection has been localized to the appropriate anatomic site. Two basic patterns of recurrent urinary tract infection may be recognized: relapse, in which the original organism is suppressed by antimicrobial therapy, and then reappears when the antibiotic is stopped; and reinfection, in which the original organism is eradicated by antimicrobial therapy, and recurrence is due to the introduction of a new bacterial strain from the fecal reservoir, to the vagina, and then into the bladder—thus initiating an entirely new bout of urinary tract infection. Stamey (10–13) showed one particular type of reinfection that may

be difficult to distinguish from relapsing infection is that in which the organism is eradicated from the urinary tract with therapy, but persists in the stool, then recolonizes the vagina, and subsequently reinfects the bladder. We observed that true relapses can be distinguished from this variety of reinfection by the time course of recurrence. Most relapses occur within one week after the completion of therapy; while most such reinfections occur several weeks to months later.

Turck (31, 50, 51) and his colleagues have clearly demonstrated on the basis of careful bilateral ureteral catheterization studies that the great majority of relapsing infections occur in the patients with renal bacteriuria, whereas the great majority of reinfections occur in patients with infection restricted to the bladder. They studied 131 women with 262 recurrent infections over a period of 18 months. Sixty-eight had bladder infection, and 63 renal infection. Whereas 121 of 149 recurrences in patients with renal bacteriuria were due to relapses, 89 of 113 recurrences among patients with bladder bacteriuria were due to reinfection with new organisms ($P < 0.001$). Thus, the pattern of recurrence following conventional antimicrobial therapy (relapse vs reinfection) is approximately 80% accurate in defining the anatomic site of infection.

It would then seem likely that the anatomy of the infection could be correlated with the antibiotic therapy necessary to eradicate the infection from the different sites: Turck et al (31, 50, 51) reported that patients with true relapsing infection—i.e. those with renal infection—will have a higher cure rate when treated for six weeks, as opposed to two weeks. Boutros et al (52) showed that the use of 100 ml of 0.2% neomycin to wash out the bladder eliminated infection in 20 of 36 patients (55%) with bladder bacteriuria, but not in patients with renal bacteriuria. Ronald et al (53), using the bladder washout technique to localize the site of infection, studied the responses of 104 patients to a single 500-mg dose of intramuscular kanamycin. In this study, 36 of 39 patients (92%) with lower tract infections were cured, as opposed to only 18 of the 65 patients (28%) with renal infections.

We utilized the antibody-coated bacteria (ACB) assay to study the response to antimicrobial therapy in two populations: normal ambulatory women with acute urinary tract infection (9), and a renal transplant population (54). Among 61 normal women presenting with symptoms of cystitis there was a direct correspondence between the results of the ACB assay and the response to antimicrobial therapy. In patients with infection localized to the bladder (ACB negative) a single dose of 3 grams of amoxicillin was shown to be 100% effective in eradicating the initial infecting organism, as was a conventional course of 10 days of 250 mg, 4 times a day, with the same agent. In contrast, 9 of 18 patients (50%) with upper tract infection as delineated by a positive ACB test, relapsed within one week of discon-

tinuing a full 10 days of antimicrobial therapy. When the therapeutic results achieved in this study population are thoroughly analyzed, it is apparent that the initial pathogen was eradicated from 52 of 61 patients (85%), with all the treatment failures coming from the group of patients with renal infection. These results (85% overall success rate) are comparable to those achieved in a number of studies with a large variety of antimicrobial agents in which the site of infection was not localized, which leads us to the following conclusions: To be interpretable, any study of acute urinary tract infection that attempts to evaluate the efficacy of a particular treatment program must localize the site of infection in the patients treated. Bladder infection is relatively easy to treat, and a variety of regimens ranging from single dose to conventional therapy are quite effective in eradicating such infections. Unfortunately, currently prescribed antimicrobial regimens lasting 7–14 days do not provide adequate treatment for a significant number of renal parenchymal infections (9). The initial observations on the duration of therapy necessary in patients with relapsing renal infection by Turck, Petersdorf, and their colleagues (31, 50, 51) in the 1960s appear to be confirmed by our studies, which approached the problem from a different point of view.

The importance of infection localization in guiding antimicrobial therapy is further confirmed by our studies in the more complicated renal transplant population. A retrospective review of the occurrence of urinary tract infection in our transplant population revealed that 41% (82 of 200) of these patients developed symptomatic urinary tract infection in the first three months after renal transplantation. Such infection was complicated by a 62% relapse rate and a 12% incidence of gram negative bacteremia, despite treatment for two weeks with an effective antimicrobial agent. In an attempt to improve the management of this problem, ACB assays were performed, with ACB positive patients assigned to six weeks of therapy and ACB negative patients to two weeks of therapy. Of 19 patients studied, 14 (73%) had positive ACB tests, and were treated for six weeks—infection was eradicated in 13 of 14 patients (92%). The five ACB negative patients were treated for two weeks with antibiotics—infection was eradicated in all cases. Most important of all, this approach essentially eliminated the occurrence of gram negative bacteremia in this population (54). Thus, localization of the site of infection appears to play a useful role in guiding therapy in patients with both complicated and uncomplicated urinary tract infections.

SINGLE-DOSE THERAPY OF LOWER URINARY TRACT INFECTION

From a therapeutic point of view, it is important to note that bladder infections are superficial, and that the bladder is bathed in urine in which

high concentrations of effective antimicrobial agents may be easily achieved. An analogy can be drawn to another superficial mucosal infection, uncomplicated gonorrhea, in which single-dose antibiotic therapy for antibiotic-sensitive strains of *Neisseria gonorrhoeae* results in a 95% success rate. This suggested that single-dose amoxicillin therapy, using the same 3-gram dose successfully employed in gonorrhea, would be effective for bladder infection (9). In the study summarized above (9), and in unpublished studies in which an additional 26 patients with lower urinary tract infection were successfully treated with single-dose amoxicillin therapy, this expectation has been fulfilled. Not surprisingly, other investigators have attempted to use single-dose therapy. In the mid-1960s, single doses of ultra-long-acting sulfonamides (drugs no longer marketed in the United States), with half-lives of 150–200 hours within the body, were shown to be as effective as conventional treatment regimens (55, 56). Brumfitt et al (57), reported a 56% failure rate with a single intramuscular dose of 2 grams of cephaloridine. However, Ronald et al (53), in the study mentioned above, noted the great efficacy of intramuscular kanamycin for lower urinary tract infection, and its lack of value for renal infection. Bailey & Abbott (58, 59), without doing infection-localizing studies, reported an approximately 75–80% success rate with conventional amoxicillin and trimethoprim-sulfamethoxazole therapy, as well as with single doses of 3 grams of amoxicillin and the equivalent of three double-dose tablets of trimethoprim-sulfamethoxazole. Thus, at least three currently available agents—kanamycin (53), amoxicillin (9, 58), and trimethoprim-sulfamethoxazole (59)—appear to be effective in a single-dose regimen for uncomplicated lower urinary tract infection; the greatest experience to date is with amoxicillin. What other antimicrobial agents can be used in such a treatment program, and whether such a treatment program will be applicable in children, remain to be established.

The ease and convenience of administering single-dose therapy, the decreased cost and side effects, and the better patient compliance when compared with conventional therapy are obvious. In addition, preliminary data suggest that the response to single-dose therapy may be a useful guide in patient management. All of our patients with ACB positive infection who initially received a single dose of amoxicillin (a total of 22 patients now) relapsed by culture (and, in about half the cases, relapsed symptomatically) within four days after treatment with the single dose. Given the difficulties of infection-localization techniques for the practicing physician (even with the ACB assay), a strong argument can be made for giving single-dose therapy to all patients presenting with clinical cystitis, and having them return for follow-up 3–5 days later. In the majority of patients (50–70%) who have infection restricted to the lower urinary tract, the bacteria will be eradicated and no further intervention will be necessary. In the significant minority (30–50%) with upper urinary tract infection, relapse will

have occurred, and more intensive therapy (possibly for six weeks) can be prescribed. Thus, therapy can be tailored rather easily to the site of infection. In addition, one would predict that those patients with infection superimposed on anatomical defects of the urinary tract requiring surgical correction would fail the trial of single-dose therapy. Therefore, one would have a marker for narrowing down the population that should be subjected to radiologic evaluation of the urinary tracts (9). Studies to test this hypothesis in both adults and children have been initiated, and preliminary results are consistent with this hypothesis.

Three years ago, in this same publication, Sanford (8) emphasized the importance of localization studies in prescribing antimicrobial therapy and predicted the efficacy of single-dose therapy for lower urinary tract infection. The work done since then has confirmed the wisdom of his predictions and analysis.

SUMMARY

Urinary tract infection in adult females is the most common bacterial infection affecting this population, with an annual infection rate involving at least 6% of women. Although recently published information suggests that urinary tract infection does not play a primary role in the pathogenesis of chronic renal failure, it may play an important secondary role. In any case, the morbidity associated with this common problem more than justifies the attempts made to improve therapy and to prevent recurrence. Two forms of recurrence are noted: relapse due to inadequate treatment of the original organism, and reinfection with a new organism—the latter is not a true treatment failure. The majority of relapses occur in patients with renal infection; the majority of therapeutic cures and reinfections occur in patients with infection restricted to the bladder. This difference in clinical behavior reflects the differing pathophysiology of the two sites of infection: renal infection is a deep medullary tissue infection; bladder infection is a superficial mucosal infection. Delineation of the site of infection on clinical grounds alone is impossible, and although a variety of ingenious techniques have been investigated, only three of these seem to provide adequate information about the site of infection: the invasive techniques of bilateral ureteral catheterization and bladder washout, which are useful only for research purposes; and the noninvasive technique of demonstrating the presence of antibody coating of bacteria in the urine sediment (positive results are equated with kidney infection, negative with bladder infection). Studied with these techniques, both single-dose and conventional antibiotic therapy appear to be virtually always effective for bladder infection, while even conventional therapy has a high failure rate for renal infection. Indeed, the

response to single-dose therapy may prove to be a useful index in choosing both the ideal antimicrobial therapy needed and the type of diagnostic evaluation indicated. In any case, the merits of both infection localization and single-dose therapy of lower urinary tract infection appear to be clearly established.

Literature Cited

1. Heptinstall, R. H. 1974. *Pathology of the Kidney*, pp. 821-927. Boston: Little, Brown. 2nd ed.
2. Heptinstall, R. H. 1976. Interstitial nephritis. *Am. J. Pathol.* 83:214-32
3. Freeman, R. B., Smith, W. M., Richardson, J. A., Flennelly, P. J., Thurm, R. H., Urner, C., Villancourt, J. A., Griep, R. J., Bromer, L. 1975. Long term therapy for chronic bacteriuria in men; USPHS Cooperative Study. *Ann. Intern. Med.* 83:133-47
4. Asscher, A. W., Sussman, M., Waters, W. E., Evans, J. A. S., Campbell, H., Evans, K. T., Williams, J. E. 1969. Asymptomatic significant bacteriuria in the non-pregnant woman: II. Response to treatment and follow-up. *Br. Med. J.* 1:804-6
5. Asscher, A. W., Sussman, M., Waters, W. E., Evans, J. A. S., Campbell, H., Evans, K. T., Williams, J. E. 1969. The clinical significance of asymptomatic bacteriuria in the non-pregnant woman. *J. Infect. Dis.* 120:17-21
6. Murray, T., Goldberg, M. 1975. Chronic interstitial nephritis: etiologic factors. *Ann. Intern. Med.* 82:453-59
7. Kass, E. H., Savage, W. Santamarina, B. A. G. 1965. The significance of bacteriuria in preventive medicine, *Progress in Pyelonephritis*, ed. E. H. Kass, pp. 3-10. Philadelphia: Davis
8. Sanford, J. P. 1976. Urinary tract symptoms and infections. *Ann. Rev. Med.* 26:485-98
9. Fang, L. S. T., Tolkoff-Rubin, N. E., Rubin, R. H. 1978. Efficacy of single dose and conventional amoxicillin therapy in urinary tract infection localized by the antibody-coated bacteria technic. *N. Engl. J. Med.* 298:413-16
10. Stamey, T. A. 1972. *Urinary Infections*. Baltimore: William & Wilkins. 293 pp.
11. Schaeffer, A. J., Stamey, T. A. 1977. Studies of introital colonization in women with recurrent urinary tract infections. IX. The role of antimicrobial therapy. *J. Urol.* 118:296-98
12. Fowler, J. E. Jr., Stamey, T. A. 1977. Studies of introital colonization in women with recurrent urinary infections. VII. The role of bacterial adherence. *J. Urol.* 117:472-76
13. Stamey, T. A., Wehner, N., Mihara, G., Condy, M. 1978. The immunologic basis of recurrent bacteriuria: role of cervicovaginal bleeding in enterobacterial colonization of the introital mucosa. *Medicine* 57:47-56
14. Bran, J. L., Levison, M. E., Kaye, D. 1972. Entrance of bacteria into the female urinary bladder. *N. Engl. J. Med.* 286:626-29
15. Buckley, R. M. Jr., McGuekin, M., MacGregor, R. R. 1978. Urine bacterial counts after sexual intercourse. *N. Engl. J. Med.* 298:321-24
16. Rocha, H. 1972. Pathogenesis and clinical manifestations of urinary tract infection. In *Urinary Tract Infection and Its Management*, ed. D. Kay, pp. 6-27. St. Louis: Mosby
17. Gallagher, D. J. A., Montgomerie, J. Z., North, J. D. K. 1965. Acute infections of the urinary tract and the ureteral syndrome in general practice. *Br. Med. J.* 1:622-26
18. Fairley, K. F., Carson, N. E., Gutch, R. C., Leighton, P., Grounds, A. D., Laird, E. C., McCallum, P. H. G., Sleeman, R. L., O'Keefe, C. M. 1971. Site of infection in acute urinary tract infection in general practice. *Lancet* 2:615-18
19. Stamey, T. A., Govan, D. E., Palmer, J. M. 1965. The localization and treatment of urinary tract infections: the role of bactericidal urine levels as opposed to serum levels. *Medicine* 44:1-36
20. Fairley, K. F., Bond, A. G., Brown, R. B., Habersberger, P. 1967. Simple test to determine the site of urinary tract infection. *Lancet* 2:427-28
21. Turck, M. 1975. Localization of the site of recurrent urinary tract infection in women. *Urol. Clin. North Am.* 2: 433-41
22. Kass, E. H., Zinner, S. H. 1969. Bacteriuria and renal disease. *J. Infect. Dis.* 120:27-46
23. Kaitz, A. L., London, A. M. 1964. Osmolar urinary concentrating ability and pyelonephritis in hospitalized patients. *Am. J. Med. Sci.* 248:7-15

24. Kaitz, A. L. 1961. Urinary concentrating ability in pregnant women with asymptomatic bacteriuria. *J. Clin. Invest.* 40:1331-38
25. Norden, C. W., Levy, P. S., Kass, E. H. 1970. Predictive effect of urinary concentrating ability and hemagglutinating antibody titer upon response to antimicrobial therapy in bacteriuria of pregnancy. *J. Infect. Dis.* 121:588-596
26. Clark, H., Ronald, A. R., Cutler, R. E., Turck, M. 1969. The correlation between site of infection and maximal concentrating ability in bacteriuria. *J. Infect. Dis.* 120:47-53
27. Ronald, A. R., Cutler, R. E., Turck, M. 1969. Effect of bacteriuria on the renal concentrating mechanism. *Ann. Intern. Med.* 70:723-33
28. Wacker, W., Dorfman, L. 1962. Urinary lactic dehydrogenase activity. *J. Am. Med. Assoc.* 181:148-54
29. Bank, N., Baline, S. H. 1965. Urinary beta-glucuronidase activity in patients with urinary tract infection. *N. Engl. J. Med.* 272:70-75
30. Ronald, A. R., Silverblatt, F., Clark, H., Cutler, R. E., Turck, M. 1971. Failure of urinary beta-glucuronidase activity to localize the site of urinary tract infection. *Appl. Microbiol.* 21:990-92
31. Turck, M. 1979. Importance of localization studies of urinary tract infection in women. In *Symposium on Pyelonephritis*, ed. E. H. Kass, W. Brumfitt. Univ. Chicago Press. In press
32. Jodal, U., Lindberg, U., Lincoln, K. 1975. Level diagnosis of symptomatic urinary tract infections in childhood. *Acta Paediatr. Scand.* 64:201-8
33. Jodal, U., Hanson, L. A. 1976. Sequential determination of C-reactive protein in acute childhood pyelonephritis. *Acta Paediatr. Scand.* 65:319-22
34. Sabel, K. G., Hanson, L. A. 1974. The clinical usefulness of C-reactive protein (CRP) determinations in bacterial meningitis and septicemia in infancy. *Acta Paediatr. Scand.* 63:381-88
35. Holmgren, J., Smith, J. W. 1975. Immunological aspects of urinary tract infections. *Progr. Allergy* 18:289-352
36. Cotran, R. S. 1963. Retrograde *Proteus* pyelonephritis in rats. Localization of antigen and antibody in treated sterile pyelonephritic kidneys. *J. Exp. Med.* 117:813-21
37. Lehrman, J. D., Smith, J. W., Miller, T. E., Barnett, J. A., Sanford, J. P. 1968. Local immune response in experimental pyelonephritis. *J. Clin. Invest.* 47:2541-2550
38. Percival, A., Brumfitt, W., DeLouvois, J. 1964. Serum antibody levels as an indication of clinically inapparent pyelonephritis. *Lancet* 2:1027-33
39. Clark, H., Ronald, A. R., Turck, M. 1971. Serum antibody response in renal versus bladder bacteriuria. *J. Infect. Dis.* 123:539-43
40. Thomas, V., Shelokov, A., Forland, M. 1974. Antibody-coated bacteria in the urine and the site of urinary tract infection. *N. Engl. J. Med.* 290:588-90
41. Jones, S. R., Smith, J. W., Sanford, J. P. 1974. Localization of urinary tract infections by detection of antibody-coated bacteria in urine sediment. *N. Engl. J. Med.* 290:591-93
42. Montplaisir, S., Cote, P. A., Martineau, B., Roche, A. J., Mongeau, J.-G., Robitaille, P. 1976. Localisation du site de l'infection urinaire chez l'enfant par la recherche de bacteries recouvertes d'anticorps. *Can. Med. Assoc. J.* 115:1096-99
43. Thomas, V. L., Harris, R. E., Gilstrap, L. C. III. 1975. Antibody-coated bacteria in the urine of obstetrical patients with acute pyelonephritis. *J. Infect. Dis.* 131: Suppl. pp. S57-S61
44. Montplaisir, S., Courteau, C., Roche, A. J. 1977. Antibody-coated bacteria in contaminated urine specimens. *N. Engl. J. Med.* 296:758
45. Ryan, R. W., Kowalski, I., Tilton, R. C. 1977. *Mechanisms of the antibody-coated bacteria test (Abstr. C2)*. Presented at 77th Ann. Meet. Am. Soc. Microbiol., New Orleans
46. Braude, R., Block, C. 1977. Proteinuria and antibody-coated bacteria in the urine. *N. Engl. J. Med.* 297:617-18
47. Jones, S. R. 1974. Prostatitis as a cause of antibody-coated bacteria in the urine. *N. Engl. J. Med.* 291:365
48. Smith, J. W., Jones, S. R., Kaijser, B. 1977. Significance of antibody-coated bacteria in urinary sediment in experimental pyelonephritis. *J. Infect. Dis.* 135:577-81
49. Hellerstein, S., Kennedy, E., Nussbaum, L., Rice, K. 1978. Localization of the site of urinary tract infections by means of antibody-coated bacteria in the urinary sediments. *J. Pediatr.* 92:188-93
50. Turck, M., Anderson, K. N., Petersdorf, R. G. 1966. Relapse and reinfection in chronic bacteriuria. *N. Engl. J. Med.* 275:70-73
51. Turck, M., Ronald, A. R., Petersdorf, R. G. 1968. Relapse and reinfection in chronic bacteriuria. II. The correlation

- between site of infection and pattern of recurrence in chronic bacteriuria. *N. Engl. J. Med.* 278:422-27
52. Boutros, P., Mourtada, H., Ronald, A. R. 1972. Localization of urinary infection. *Am. J. Obstet. Gynecol.* 112: 379-81
 53. Ronald, A. R., Boutros, P., Mourtada, H. 1976. Bacteriuria localization and response to single-dose therapy in women. *J. Am. Med. Assoc.* 235: 1854-56
 54. Rubin, R. H., Fang, L. S. T., Cosimi, A. B., Herrin, J. T., Varga, P. A., Russell, P. S., Tolkoff-Rubin, N. E. 1978. The usefulness of the antibody-coated bacteria assay in the management of urinary tract infection in the renal transplant patient. *Transplantation*. In press
 55. Gruneberg, R. N., Brumfitt, W. 1967. Single-dose treatment of acute urinary tract infection: a controlled trial. *Br. Med. J.* 3:649-51
 56. Williams, J. D., Smith, E. K. 1970. Single-dose therapy with streptomycin and sulfametopyrazine for bacteriuria during pregnancy. *Br. Med. J.* 4:651-53
 57. Brumfitt, W., Faiers, M. C., Franklin, I. N. S. 1970. The treatment of urinary infection by means of a single dose of cephaloridine. *Postgrad. Med. J.* 46: Suppl. pp. 65-69
 58. Bailey, R. R., Abbott, G. D. 1977. Treatment of urinary tract infection with a single dose of amoxycillin. *Nephron* 18:316-20
 59. Bailey, R. R., Abbott, G. D. 1978. Treatment of urinary tract infection with a single dose of trimethoprim-sulfamethoxazole. *Can. Med. Assoc. J.* 118:551-52

BIG ACTH AND BRONCHOGENIC CARCINOMA

♦7319

Rosalyn S. Yalow, Ph.D.

Solomon A. Berson Research Laboratory, Veterans Administration Hospital, Bronx, New York 10468; and The Mount Sinai School of Medicine, CUNY, New York, New York 10029

INTRODUCTION

It is now commonly accepted that a variety of nonendocrine tumors are capable of elaborating one or more humoral substances that resemble well-known, well-characterized, biologically active hormones or their prohormone precursors or metabolic products (see 1-3 for reviews). These observations have promoted considerable interest in the potential of mass screening programs for early detection of malignancies. These programs would be based on measurement, generally by radioimmunoassay (RIA), of elevated concentrations of one or another of these hormonal forms in the circulating blood. Or such RIAs might provide objective measures of tumor responsiveness to various therapeutic procedures. This review considers the potential role of ACTH determinations in the diagnosis and management of patients with bronchogenic carcinoma since it is among such patients that ectopic Cushing's syndrome occurs most frequently (1).

CHARACTERIZATION OF ACTH IN TISSUE AND PLASMA

In our early studies (4, 5) using RIA of plasma ACTH we did not notice any discrepancy between hormonal concentration and the clinical state or responsiveness to modulators of ACTH secretion. Nonetheless we subsequently observed (6) that in patients with pituitary Cushing's syndrome, plasma immunoreactive ACTH was generally considerably lower than in patients with ectopic Cushing's syndrome. Similar observations were re-

cently reported by Rees (7). The differences between the two groups of patients is now known to be related to the differences in the hormonal forms of immunoreactive ACTH in the circulation.

Convincing evidence for the heterogeneity of other peptide hormones led us to investigate the nature of ACTH in plasma and tissue. We detected and characterized a new form of ACTH ("big ACTH"), which is larger and more acidic than the usual 1-39 peptide (8, 9). The biologic activity of big ACTH was less than 4% of its immunologic activity (10). Although the usual 1-39 peptide generally predominates in the pituitary and in the plasma after pituitary stimulation, big ACTH is usually more prominent in plasma and tumor tissue of patients with ectopic hormone production (11). Odell et al (12) confirmed our findings with the demonstration that in a group of 38 patients with carcinoma and no clinical evidence of ectopic Cushing's syndrome immunoreactive ACTH is elevated, but biologically active ACTH, as determined in a receptor assay, is normal. Furthermore, Sephadex gel filtration of some of the plasmas from these patients revealed the presence of big ACTH. Orth et al (13) initially reported that extracts of the tumors in patients with ectopic ACTH syndrome contained an ACTH similar to pituitary ACTH and both N-terminal and C-terminal ACTH fragments. More recently this same group, using a different antiserum, confirmed (14) that big ACTH predominated in extracts of tumors from two patients with ectopic Cushing's syndrome and that no immunoreactive fragments smaller than the 1-39 peptide were noted. It would certainly have been of interest to determine the Sephadex elution profiles of these same extracts with the antisera they had previously employed to determine whether these new extracts contained ACTH fragments.

It is now generally agreed that big ACTH of low biological activity is frequently a major component of plasma and tumor ACTH in patients with ectopic Cushing's syndrome. This resembles the findings that proinsulin is more prominent in tumor extracts and plasma of patients with islet cell tumors than in normal pancreatic tissue or plasma of other subjects (15, 16). These observations suggest that in tumors there may be a defect in the conversion of prohormone to active hormone. This defect may arise from a qualitative or quantitative deficiency in the converting enzyme, a defect in the "packaging" of the enzyme together with the prohormone, which would inhibit enzymatic conversion, or the presence of an enzyme-inhibiting factor in tumor tissue.

IMMUNOREACTIVE ACTH IN TISSUE EXTRACTS

Observation of elevated plasma levels of ACTH due to ectopic hormone production is probably limited by the percentage of cases in which the tumor contains hormone in excess of that found in normal, nonendocrine,

nonmalignant tissue (11). It is, of course, possible for a tumor to secrete but be devoid of a hormone because of rapid turnover, but this is not generally so for peptide-hormone-secreting tumors.

In our preliminary report (11) we were indeed surprised to detect ACTH in three bronchial adenomas and in 28 of 29 tissue samples (concentrations 4–55 ng/g wet weight tissue) from carcinoma primary to or metastatic from the lung but not from tumors metastatic to the lung. In this small sampling we could not determine whether there was a relation between tumor type and hormone concentration since some were surgical specimens and others were obtained postmortem, often many hours after death. We subsequently noted (6) that although ACTH is found in all tumor types, the concentrations in adenocarcinoma are generally significantly lower than in epidermoid carcinoma (Figure 1). Since these 44 specimens were obtained at surgical resection, patients with small-cell undifferentiated ("oat" cell) carcinoma were not included in the group.

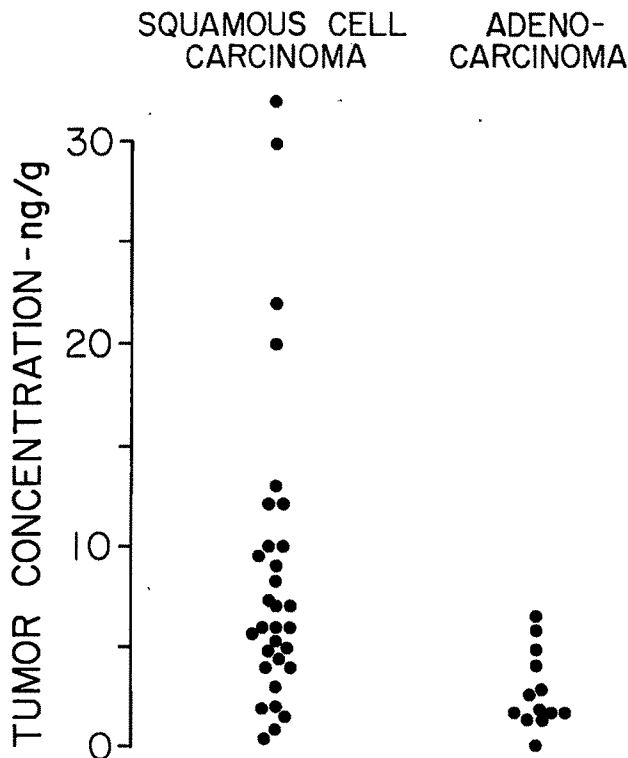


Figure 1 Range of concentrations of ACTH found in tumors of patients with squamous cell carcinoma and adenocarcinoma of the lung but without Cushing's syndrome. (Reproduced from Ref. 6)

ACTH is not detectable (< 1 ng/g wet weight tissue) in normal lung of man, dog, or rabbit. However, lung tissue from a "smoking" dog without invasive carcinoma but showing atypical histologic changes, including basal cell hyperplasia and atypical proliferation of alveolar lining cells, contained measurable levels of ACTH, primarily in the big form: lung tissue from three other "smoking" dogs with no significant histologic changes did not (11). (These tissues were obtained from Dr. Oscar Auerbach, East Orange Veterans Administration Hospital.)

In another group of 19 patients, sampled at autopsy, immunoreactive ACTH was not detectable in normal areas of the lung. It ranged up to 34 ng/g in extracts of lung tumors; the highest values were in two patients with small-cell carcinoma. In these cases one cannot determine the possible losses in tissue ACTH between death and the time of sampling. ACTH was not detectable in a lymphoma or a renal cell carcinoma metastatic to the lung. Unexpected was the findings of levels up to 16 ng/g in areas of squamous metaplasia and suppurative bronchitis or pneumonia (R. Yalow et al, previously unpublished observations).

Thus, although ACTH is not detectable in normal lung tissue, it is found in some pathologic lung tissue even in the absence of invasive carcinoma. Important questions yet to be answered include: What is the insult to the lung and what is the extent of injury necessary to initiate elaboration of ACTH? Would bronchitis or pneumonia per se without the concomitant squamous metaplasia suffice?

The need to determine at what stage in the progression from normalcy to invasive carcinoma in the lung parenchyma of smokers the ability to elaborate ACTH first occurs is evident from the classic work of Auerbach and associates (17). They noted on the basis of 20,000 sections from the tracheobronchial tree of over 400 autopsied cases that the percentage of cases in which carcinoma in situ was found in one or more sections was not very much different between very heavy smokers who died from lung cancer (83%) and those who did not (75%). Other changes in the lung parenchyma were also quite similar between these groups (17).

IMMUNOREACTIVE ACTH IN PLASMA

Our observations that virtually all primary lung tumors contain ACTH and our initial demonstration that 21 of 24 patients with carcinoma of the lung untreated before sampling had afternoon plasma ACTH concentrations above 150 pg/ml, the upper limit achieved by only 6% of the control group, suggested the potential usefulness of plasma ACTH as a biologic marker for detection of lung cancer (11, 18). However, in a subsequent study (19) plasma ACTH was found to be elevated above 150 pg/ml in only 24 of 36 patients whose tumors were considered suitable for surgical resection. It

was elevated in only 66 of 136 patients randomized for chemotherapeutic protocols by the Veterans Administration Lung Group (6). Those with extensive disease appeared to exhibit elevated levels (59 of 107 cases) more frequently than those with limited disease, i.e. where the tumor was thought to be confined to one hemithorax, (11 of 29). In patients with extensive disease, those with adenocarcinoma exhibited elevated values less frequently (7 of 23) than those with other tumor types (52 of 84). This is consistent with the low concentrations of ACTH we found in extracts of adenocarcinoma compared to those of squamous cell carcinoma (Figure 1).

Since the minimal concentration of ACTH detectable in tumors is only about one millionth that of the pituitary (11), the finding of false negatives is not surprising. If elevated values of plasma ACTH are derived from tumor ACTH then the tumor must be large, have a higher than minimally detectable hormone concentration, or be more active in secretion of the hormone than the pituitary. Since the metabolic turnover of the big ACTH may be markedly prolonged compared to the turnover of the 1-39 peptide, the relative plasma concentration of the big form compared to the usual form would be amplified by a factor equal to the relative turnover times. However, secretion of big ACTH is not suppressed by intravenous hydrocortisone under the same conditions as is the 1-39 peptide (11). Therefore we do not know its disappearance rate and cannot determine this amplification factor.

The greatest limitation of plasma ACTH assay for mass screening, as with all other "tumor marker" assays, is the potential for false positives, i.e. elevation of marker level in the absence of carcinoma. On occasion, sporadic release of pituitary ACTH can result in elevation of plasma ACTH even in normal subjects. The number of false positives on this basis can be reduced by multiple sampling in the afternoon when ACTH secretion is diminished, by the use of dexamethasone to suppress pituitary ACTH secretion, or by fractionation on Sephadex to ascertain the particular form of circulating ACTH. Thus distinction between a control group of nonsmokers and patients with lung carcinoma can generally be made. However, for diagnostic purposes the significant distinction to be made is between the very large group of heavy smokers without carcinoma and the very small group with carcinoma. Our earlier observations that even in the absence of proven carcinoma about one third of 36 patients with chronic obstructive pulmonary disease have elevated plasma ACTH (18) suggested that a significant fraction of false positives might be found among heavy smokers. Furthermore lung tissues from a "smoking" dog with atypia (11) and from humans with squamous metaplasia, particularly with superimposed inflammatory processes, contain ACTH. Thus the ability to elaborate ACTH is manifest in injured but nonmalignant tissue.

of recurrence or exacerbation is probably dependent on the degree to which nontumorous tissue contributes to the circulating ACTH. We earlier noted (19) that in 18 of 23 patients plasma ACTH in samples taken within three months of surgical resection were lower than levels measured before surgery. We extended these studies to a group of 50 patients, 19 of whom had plasma ACTH levels greater than 300 pg/ml preoperatively. Eleven of these 19 patients had significant decreases in their circulating levels within three months of surgery. Of these, all but one was diagnosed as epidermoid carcinoma and five had pneumonectomies. Eight others had no decrease: four of these had epidermoid carcinoma and four had pneumonectomies (R. Yalow et al, previously unpublished observations). As yet no clear picture of predictability has emerged.

CONCLUSIONS

These studies suggest that determination of plasma ACTH in a mass screening program for early detection of lung carcinoma is of limited value, primarily because the levels overlap considerably between those with invasive carcinoma and those at high risk, i.e. heavy smokers (particularly those with chronic obstructive pulmonary disease). Data are as yet insufficient to determine whether low values before or marked decreases in plasma ACTH after surgery have meaningful prognostic value for remission or long-term survival. On occasion we observe a dramatic decrease in plasma ACTH on some chemotherapeutic schedules, but this does not seem to occur with sufficient regularity to serve as an objective measure of chemotherapeutic or radiotherapeutic efficacy. It seems more certain that the primary role of measurement of plasma and tissue ACTH in carcinoma of the lung will be to aid our understanding of the disease, its origin and its treatment.

ACKNOWLEDGMENT

This work was supported by the Medical Research Service of the Veterans Administration and by an Interagency Agreement between the Veterans Administration and the National Cancer Institute. We are grateful to the VA Surgical Adjuvant and the VA Lung Groups for supplying many of the samples reported on in this study.

Literature Cited

1. Liddle, G. W., Nicholson, W. E., Island, D. P., Orth, D. N., Abe, K., Lowder, S. C. 1969. Clinical and laboratory studies of ectopic humoral syndromes. *Recent Progr. Horm. Res.* 25:283-314
2. Rees, L. H. 1975. The biosynthesis of hormones by non-endocrine tumours—A review. *J. Endocrinol.* 67:143-75
3. Odell, W. D., Wolfson, A. R. 1978. Humoral syndromes associated with cancer. *Ann. Rev. Med.* 29:379-406
4. Yalow, R. S., Glick, S. M., Roth, J., Berson, S. A. 1964. Radioimmunoassay of human plasma ACTH. *J. Clin. Endocrinol. Metab.* 24:1219-25
5. Berson, S. A., Yalow, R. S. 1968. Radio-immunoassay of ACTH in plasma. *J. Clin. Invest.* 47:2725-51
6. Yalow, R. S. 1978. Ectopic ACTH in carcinoma of the lung. *Prog. Cancer Res. Ther.* 11:209-16
7. Rees, L. H. 1977. ACTH, lipotrophin and MSH in health and disease. *Clin. Endocrinol. Metab.* 6:137-53
8. Yalow, R. S., Berson, S. A. 1971. Size heterogeneity of immunoreactive human ACTH in plasma and in extracts of pituitary glands and ACTH-producing thymoma. *Biochem. Biophys. Res. Commun.* 44:439-45
9. Yalow, R. S., Berson, S. A. 1973. Characteristics of "Big ACTH" in human plasma and pituitary extracts. *J. Clin. Endocrinol. Metab.* 36:415-23
10. Gewirtz, G., Schneider, B., Krieger, D. T., Yalow, R. S. 1973. Big ACTH: Conversion to biologically active ACTH by trypsin. *J. Clin. Endocrinol. Metab.* 38:227-30
11. Gewirtz, G., Yalow, R. S. 1974. Ectopic ACTH production in carcinoma of the lung. *J. Clin. Invest.* 53:1022-32
12. Odell, W., Wolfson, A., Yoshimoto, Y., Weitzman, R., Fisher, D., Hirose, F. 1977. Ectopic peptide synthesis: A universal concomitant of neoplasia. *Trans. Assoc. Am. Physicians* 40:204-25
13. Orth, D. N., Nicholson, W. E., Mitchell, W. M., Island, D. P., Liddle, G. W. 1973. Biologic and immunologic characterization and physical separation of ACTH and ACTH fragments in the ectopic ACTH syndrome. *J. Clin. Invest.* 52:1756-69
14. Orth, D. N., Nicholson, W. E. 1977. Different molecular forms of ACTH. *Ann. NY Acad. Sci.* 297:27-48
15. Goldsmith, S. J., Yalow, R. S., Berson, S. A. 1969. Significance of human plasma insulin Sephadex fractions. *Diabetes* 18:834-39
16. Yalow, R. S., Berson, S. A. 1971. Fundamental principles of radioimmunoassay techniques in measurement of hormones. *Recent Adv. Endocrinol.* 238:16-34
17. Auerbach, O., Stout, A. P., Hammond, E. C., Garfinkel, L. 1961. Changes in bronchial epithelium in relation to cigarette smoking and in relation to lung cancer. *N. Engl. J. Med.* 265:253-67
18. Ayyazian, L. F., Schneider, B., Gewirtz, G., Yalow, R. S. 1975. Ectopic production of big ACTH in carcinoma of the lung: Its clinical usefulness as a biologic marker. *Am. Rev. Respir. Dis.* 3:279-87
19. Yalow, R. S. 1977. Ectopic ACTH in carcinoma of the lung. *Endocrinology, Vol. I, Proc. 5th Int. Congr. Endocrinol. Hamburg, July 18-24, 1976.* Amsterdam: Excerpta Med. 403:419-24

PROLACTIN AND THE DIAGNOSIS OF PITUITARY TUMORS

♦7320

William B. Malarkey, M.D.

Division of Endocrinology and Metabolism, The Ohio State University Hospital,
Columbus, Ohio 43210

INTRODUCTION

Galactorrhea associated with a pituitary tumor was described by Forbes et al in 1954 (1). Ten years later Canfield & Bates (2) presented bioassay evidence of excessive lactogenic activity in the sera of patients with pituitary tumors. Prolactin, however, had not yet been shown to be present in the pituitary of man, and some investigators attributed serum lactogenic activity to growth hormone (GH). A report by Peake et al (3) that extracts of a pituitary tumor from a lactating woman to whom anti-GH sera had been added still had bioassayable lactogenic activity suggested the presence of prolactin (PRL) in this tumor. Finally the development of specific radioimmunoassays for human PRL (4-6) clearly documented the existence of a human PRL molecule separate from human GH. It was then possible to demonstrate that the galactorrhea of many patients with pituitary tumors was associated with elevated serum PRL levels (7-9). Excessive PRL secretion can now be added to the list of excessive hormonal syndromes associated with pituitary tumors. This list presently includes excessive GH secretion (acromegaly), abnormal ACTH production (Cushing's disease), thyroid-stimulating hormone (TSH) hypersecretion (10), excessive secretion of gonadotropins (11), and increased subunit secretion of glycoprotein hormones (12).

Excessive PRL secretion from a pituitary tumor, a prolactinoma, is the most common of the pituitary excess syndromes. Approximately two thirds of patients with pituitary tumors not associated with acromegaly or Cushing's syndrome will secrete excessive PRL (13). These pituitary tumors have been referred to as chromophobe adenomas because routine staining failed

to demonstrate any increased hormonal content. Several groups, however, have shown that different handling of the surgical specimens and the use of more specific staining techniques provide evidence for increased PRL production by many of these pituitary tumors (9, 14). At the ultrastructural level there are certain characteristic findings of these neoplasms such as extensive development of the rough endoplasmic reticulum and golgi apparatus with misplaced exocytosis of secretory granules away from the capillary site of the cell (15, 16). These anatomic findings are seen in patients with elevated serum PRL levels whether or not galactorrhea is present.

The excessive PRL secretion of these tumors may on rare occasion be associated with excessive follicle-stimulating hormone (FSH) (17) and TSH (18) secretion. In addition many patients with excessive GH secretion and acromegaly will have an associated hypersecretion of PRL (19). This latter event usually reflects the secretion from two separate cell lines within the adenoma (20).

Over the past five years we have seen approximately 34 patients with a definite diagnosis of a pituitary tumor not associated with acromegaly or Cushing's disease in whom serum prolactin determinations were performed. Twenty-two (65%) of these patients had elevated serum prolactin levels with a range of 15–4800 ng/ml in 15 males and 24–480 ng/ml in 19 women (a normal PRL level in females is 3–20 ng/ml, and 3–15 ng/ml is normal in males). Amenorrhea or irregular menses were common (92%) in the premenopausal women with pituitary tumors and hyperprolactinemia, but were also frequent (80%) in women with pituitary tumors and normal prolactin levels. Galactorrhea was found in 61% of the women with hyperprolactinemia and pituitary tumors. In addition, several of the women with pituitary tumors and normal serum prolactin levels were lactating. Abnormal visual fields were found in 58% of these patients and this finding was not related to the presence of hyperprolactinemia. These clinical findings are similar to those reported by Kleinberg et al (19) in 48 pituitary tumor patients.

DIAGNOSTIC CONSIDERATIONS

A serum PRL determination is helpful in evaluating many patients that the clinician suspects of having a pituitary tumor. Its value, however, is greatly influenced by the manner in which the patient presents. For example, consider the patient who presents to the ophthalmologist or neurosurgeon with a visual field defect and an enlarged sella turcica. In this patient a CAT scan may reveal the presence of a suprasellar mass lesion. In this situation a serum PRL determination is of little diagnostic value as the presence of a pituitary tumor is almost certain.

In contrast to this setting is the more frequent clinical presentation of a patient with headaches who has a slightly enlarged sella turcica on skull X ray or the patient with infertility and amenorrhea with a normal skull X ray. In these latter two situations the presence of a pituitary tumor has to be differentiated from the empty sella syndrome, the various causes of secondary amenorrhea, and idiopathic hyperprolactinemia. A serum PRL determination is of great benefit in this diagnostic setting.

Differential Diagnosis of Hyperprolactinemia

There are numerous physiologic, pharmacologic, and pathologic situations that can produce hyperprolactinemia (Table 1). The proper interpretation of an elevated serum PRL level in the patient suspected of having a pituitary tumor requires knowledge of these items.

Prolactin secretion is probably regulated by hypothalamic inhibiting and releasing factor(s). Most pituitary hormones are primarily regulated by releasing factor activity, which contrasts with PRL regulation where inhibitory activity is dominant. Hence, disease of the hypothalamic-pituitary unit will tend to produce diminished secretion of various anterior pituitary hormones except for PRL, which will frequently be increased. Catecholamines also have an important role in inhibiting PRL secretion (21). Therefore drugs or diseases that block adrenergic or dopaminergic receptors of the CNS may lead to elevated serum PRL concentrations.

When other causes of an elevated serum PRL level have been excluded, a pituitary tumor must be considered strongly. The absolute level of serum PRL is helpful in the differential diagnosis. A serum PRL level greater than 200 ng/ml strongly suggests the presence of a pituitary tumor. A serum level above 300 ng/ml is probably pathognomonic of a prolactinoma (19). Several reports suggested that a relationship exists between the size of a pituitary tumor and the magnitude of the elevation in serum PRL concentration (9, 22). Most pituitary tumors, however, are associated with serum

Table 1 Causes of increased concentrations of prolactin in serum

Physiological	Pathological		Pharmacologic
Pregnancy	Pituitary tumors	Adrenal insufficiency	Various drugs (e.g.
Stress (e.g. surgery)	Pituitary stalk section	Ectopic production	chlorpromazine)
Sleep	Hypothalamic disorders, e.g. sarcoid infiltration	by tumors	Oral contraceptives
Suckling	Chiari-Frommel syndrome	Chest wall trauma	Estrogens
Exercise		Hypogonadism	Thyrotropin-releasing hormone
Sexual intercourse	Renal failure	Idiopathic precocious puberty	Insulin-induced hypoglycemia
	Hypothyroidism	Delayed puberty	Arginine
	Nelson's syndrome	Idiopathic	

PRL levels much lower than 200 ng/ml and are in the range of values found in patients with idiopathic or oral contraceptive-induced hyperprolactinemia.

Galactorrhea and Prolactin

Patients with hyperprolactinemia often have amenorrhea and/or galactorrhea. The lesion producing the elevation in serum PRL may lead to disordered gonadotropin secretion and subsequent amenorrhea. In addition the elevated serum PRL levels may produce a menstrual defect via a hypothalamic and/or a direct ovarian effect. The galactorrhea in patients with abnormal menses is usually related to the effect of increased serum PRL levels on breast milk synthesis. Many patients, however, with elevated serum PRL levels will not have galactorrhea. This observation suggests the importance of other endocrine factors as well as the sensitivity of the individual's breast tissue in the development of galactorrhea. This latter point is demonstrated by the women with nonpuerperal lactation associated with normal serum PRL levels and no associated endocrinopathy (23).

Prolactin Function Tests

SUPPRESSIVE AGENTS Numerous investigators have developed various PRL function tests in an attempt to differentiate patients with pituitary tumor from those with other forms of hyperprolactinemia.

L-dopa (7, 8) and various dopaminergic agents such as bromocryptine (24) acutely suppress serum PRL levels in man. It was hoped that PRL secretion in patients with pituitary tumors might be less responsive to these agents. It is now apparent, however, that similar suppression of serum PRL concentrations following the administration of these agents can occur in both normal individuals and in patients with pituitary tumors.

It was reported that water loading could suppress serum PRL levels in normal man and that patients with pituitary tumors were less sensitive to this inhibitory influence (25). These observations, however, have not been confirmed in normal man or in patients with pituitary tumors (26).

Reports that pyridoxine (vitamin B₆) could decrease puerperal and nonpuerperal lactation as well as diminish elevated serum PRL levels generated enthusiasm for its use as a function test (27). It was suggested that pyridoxine, which is a cofactor in the conversion of dopa to dopamine, was exerting its effect through increased dopaminergic tone. Other investigators, however, have not confirmed these observations and therefore pyridoxine infusions are not being used in the differential diagnosis of hyperprolactinemic states (28, 29). Hence catechol inhibition of PRL secretion, as well as the water loading and pyridoxine infusion tests, have been of limited value in the evaluation of a patient with a suspected pituitary tumor.

STIMULATING AGENTS A series of PRL function tests have also been developed that depend on pharmacologic stimulation of PRL secretion. The observation that 25 mg of chlorpromazine IM would increase serum PRL concentrations at least twofold above basal levels has served as the basis for a commonly used PRL function test (8, 30). This test, however, is abnormal in the presence of most disorders of PRL secretion and therefore it has been of limited value in the patient with a suspected pituitary tumor.

Thyroid-releasing hormone (TRH) has been shown to acutely increase serum PRL levels as well as serum TSH concentrations, and patients with pituitary tumors often fail to normally double their serum PRL concentrations following a TRH infusion (19). The problem with the interpretation of the TRH test, however, is related to the frequently abnormal TRH tests in patients with idiopathic hyperprolactinemias (19), as well as normal responses in some patients with pituitary tumors (31).

Thus the stimulatory, like the suppressive, PRL function tests also suffer from a lack of specificity and offer little diagnostic discrimination in the evaluation of the patient with a suspected pituitary tumor.

PHYSIOLOGIC STUDY

Serum Prolactin secretion can also be evaluated by frequent sampling through an indwelling catheter over a 24-hour interval. Patients with pituitary tumors and normal or elevated serum PRL levels have a blunting of the normal nocturnal augmentation of PRL secretion (32). Although this blunting of nocturnal PRL secretion is almost always found in patients with documented pituitary tumors, it may also be found in other conditions such as breast cancer (33), Cushing's syndrome (34), and idiopathic hyperprolactinemia (23). This finding of disordered nocturnal PRL secretion therefore should be interpreted with caution in the patient suspected of having a pituitary tumor.

Cerebrospinal fluid Prolactin is present in the cerebrospinal fluid (CSF) of man and is found in increased concentrations in the CSF of patients with pituitary tumors (35-37). It is possible that PRL enters the CSF via passive diffusion from serum or via retrograde transport from the pituitary.

The elevated CSF PRL levels found in patients with pituitary tumors are usually associated with increased serum PRL levels. Therefore in this setting they are of limited diagnostic value. A measurement of CSF PRL level may be of value, however, in the differential diagnosis of the patient with a large sella turcica and normal serum PRL levels. The two entities that have to be considered in this clinical setting are the empty sella syndrome and a pituitary tumor. The primary empty sella syndrome occurs predomi-

nately in middle-aged females with headaches and mild hypertension (38). They usually have normal endocrine function tests, even though they have a marked decrease in pituitary tissue volume. This latter finding is documented by the presence of air in the sella turcica during pneumoencephalography. Subjects with the empty sella syndrome have both normal serum and CSF PRL concentrations, whereas patients with pituitary tumors and normal serum PRL levels may have elevated CSF PRL levels (Figure 1) (35). Thus, if the physical examination, endocrine data and sella turcica appearance on skull X ray are compatible with the empty sella syndrome, a normal CSF PRL determination would further support this diagnosis.

Hence a single serum or CSF PRL level provides useful clinical information in the evaluation of a patient with a suspected pituitary tumor whereas PRL function testing at this time is of limited additional value.

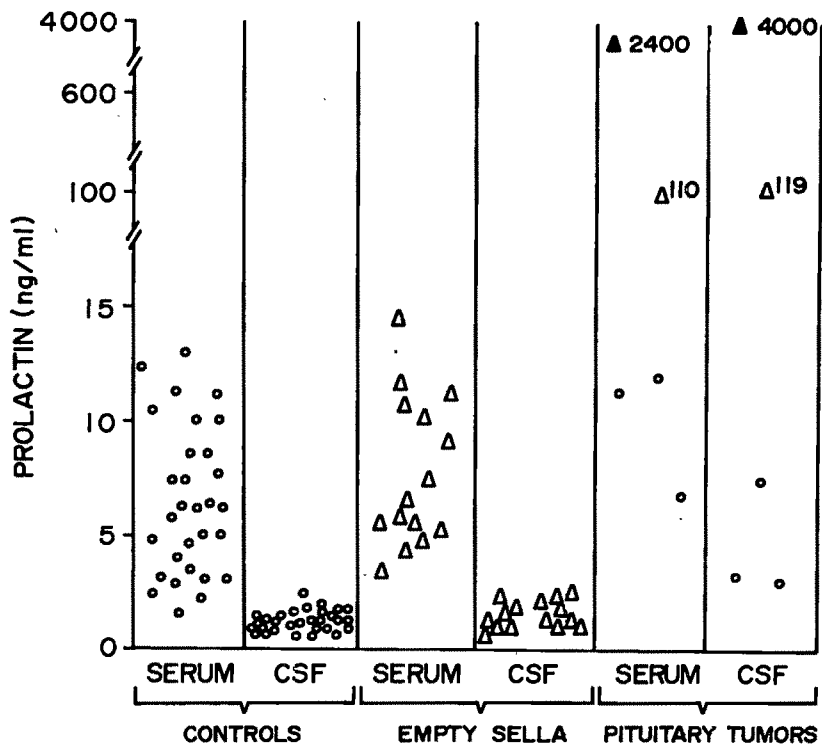


Figure 1. The serum and CSF PRL concentrations were similar in controls and patients with the empty sella syndrome. Three other subjects with chromophobe adenomas (open circles) had normal serum and elevated CSF PRL levels. A patient with acromegaly (open triangles) and another subject with a chromophobe adenoma (solid triangles) had CSF PRL concentrations that were elevated above the corresponding serum levels (35, with permission).

Interpretation of Prolactin Assay Result

Some caution, however, should be exerted in the interpretation of mildly elevated or high-normal serum PRL levels, especially at this time because of the absence of a generally accepted standard for human PRL. Hence, the immunologic potency of a PRL standard derived from a new batch of human pituitaries may differ from that which the laboratory was previously using. Hence values reported from a laboratory should reflect this phenomenon in the results reported to the clinician. Through use of various PRL standards in our laboratory we have seen differences in potency that could lead to a threefold difference in serum PRL levels unless appropriate corrections were made (unpublished observation).

Pituitary Microadenomas

In the last few years numerous publications have dealt with the issue of hyperprolactinemia and infertility in females. These women usually have galactorrhea and/or amenorrhea associated with a normal-sized sella turcica on lateral skull films. A small blister defect in the sella of these women after tomographic evaluation has been considered strongly supportive of the diagnosis of a pituitary microadenoma (39). Transsphenoidal pituitary microsurgery (40) or medical therapy (24, 40) with the dopaminergic agonist bromocryptine has permitted lowering of the serum PRL level and subsequent ovulation and successful pregnancies in these women.

This successful outcome in patients with pituitary microadenomas has encouraged investigators to develop a PRL function test that will separate the patient with idiopathic hyperprolactinemia from the individual with a pituitary microadenoma. Unfortunately there is no PRL function test presently available that can accomplish this task.

Posttherapy Prolactin Levels

A serum PRL level can also be used to evaluate patients with pituitary tumors following ablative therapy. A decline in the serum PRL level occurs quickly following a successful hypophysectomy, but the fall occurs more slowly following radiation therapy (32, 40). Since the number of patients developing pituitary tumor recurrence is small following appropriate treatment, it is too early to tell whether an increase in serum prolactin levels will be an early monitor of tumor recurrence.

PATHOPHYSIOLOGY OF THE HYPERPROLACTINEMIA

Numerous hypotheses have been offered in an attempt to explain the hyperprolactinemia associated with pituitary tumors. First, defective synthesis

and/or release of hypothalamic prolactin-inhibiting factor (PIF) or excessive release of prolactin-releasing factor could lead to hyperplasia and eventual adenoma formation. This hypothesis remains speculative without assays capable of measuring either of these factors. A recent paper, however, suggests that defective central conversion of L-dopa to dopamine is present in these patients, which may contribute to the hyperprolactinemia of these subjects (41).

Secondly, a nonfunctioning pituitary tumor could obstruct the transport of PIF to the pituitary, which would lead to an overproduction of PRL by the normal and possibly by the tumor-bearing pituitary tissue. This possibility seems unlikely in that many pituitary tumors that hypersecrete prolactin exhibit only minimal pituitary enlargement and rarely impinge on either the hypothalamus or pituitary stalk.

The third possibility is that the pituitary PRL cell has become refractory to hypothalamic control. This autonomy of PRL secretion could be related to the development of a defective receptor for PIF or a lesion distal to the binding of the hypothalamic hormone to the lactotropic receptor. It has been suggested that this is an unlikely event since dopamine, which may resemble PIF, can readily decrease human pituitary PRL release in patients with pituitary tumors (41). In contrast, studies of a rat pituitary tumor in vitro that hypersecreted PRL revealed defective dopaminergic inhibition of PRL secretion (42). This animal model, however, may not be representative of the pituitary neoplasm in man, or, alternatively, too large a dose of dopamine may have obscured a defect in the human studies. Further study of these hypotheses will be required for a clearer understanding of the pathogenesis of excessive PRL secretion in patients with pituitary tumors.

SUMMARY

The development of the radioimmunoassay for human PRL has markedly expanded our understanding of the nature of the most common type of pituitary neoplasm, the prolactinoma. Prolactin function tests (e.g. chlorpromazine and TRH stimulation) that utilize the ability of these compounds to increase serum PRL levels are frequently disordered in various hyperprolactinemic states and therefore an abnormal test is not specific for a pituitary neoplasm. Abnormal prolactin suppression tests (e.g. L-dopa or bromocriptine) are also not specific for pituitary tumors and, in fact, patients with pituitary tumors frequently have normal suppression tests. At this time a single serum PRL level is probably the most useful test of PRL secretion in a patient suspected of having a pituitary tumor. Prolactin levels above 300 ng/ml are almost pathognomonic of a pituitary neoplasm. Much effort is now being expended in an effort to develop a function test to

differentiate hyperprolactinemic women with amenorrhea and/or infertility with a pituitary microadenoma from a patient with other forms of hyperprolactinemia.

ACKNOWLEDGMENTS

This work was aided by Clinical Research Grant RR-34 from the NIH and by Grant PDT-32 from the American Cancer Society, New York, New York. In addition, I regret that space limitations have precluded reference to each major work in the area of prolactin and pituitary tumors.

Literature Cited

- Forbes, A. P., Henneman, P. H., Griswold, G. C., Albright, F. 1954. Syndrome characterized by galactorrhea, amenorrhea and low urinary FSH: comparison with acromegaly and normal lactation. *J. Clin. Endocrinol. Metabol.* 14:265-71
- Canfield, C. J., Bates, R. W. 1965. Nonpuerperal galactorrhea. *N. Engl. J. Med.* 273:897-902
- Peake, G. T., McKeel, D. W., Jarett, L., Daughaday, W. H. 1976. Ultrastructural, histologic and hormonal characterization of a prolactin-rich human pituitary tumor. *J. Clin. Endocrinol. Metab.* 29:1383-93
- Hwang, P., Gwyda, H., Friesen, H. 1971. A radioimmunoassay for human prolactin. *Proc. Natl. Acad. Sci. USA* 68:1902-6
- Jacobs, L. S., Mariz, I. K., Daughaday, W. H. 1972. A mixed heterologous radioimmunoassay for human prolactin. *J. Clin. Endocrinol. Metabol.* 34:484-90
- Sinha, N. N., Selby, F. W., Lewis, U. J., Vanderlaan, W. P. 1973. A homologous radioimmunoassay for human prolactin. *J. Clin. Endocrinol. Metab.* 36:509-16
- Malarkey, W. B., Jacobs, L. S., Daughaday, W. H. 1971. Levodopa suppression of prolactin in nonpuerperal galactorrhea. *N. Engl. J. Med.* 285:1160-63
- Friesen, H., Guyda, H., Hwang, P., Tyson, J. E., Barbeau, A. 1972. Functional evaluation of prolactin secretion: a guide to therapy. *J. Clin. Invest.* 51:706-9
- Zimmerman, E. A., Defendini, R., Frantz, A. G. 1973. Prolactin and growth hormone in patients with pituitary adenomas: a correlative study of hormone in tumor and plasma by immunoperoxidase technique and radioimmunoassay. *J. Clin. Endocrinol. Metabol.* 38:577-85
- Kourides, I. A., Ridgway, E. C., Weintraub, B. D., Bigos, S. T., Gershengorn, M. C., Maloof, F. 1977. Thyrotropin-induced hyperthyroidism: use of alpha and beta subunit levels to identify patients with pituitary tumors. *J. Clin. Endocrinol. Metabol.* 45:534-43
- Snyder, P. J., Sterling, F. H. 1976. Hypersecretion of LH and FSH by a pituitary adenoma. *J. Clin. Endocrinol. Metabol.* 42:544-50
- Korides, J. A., Weintraub, B. D., Rosen, S. W., Ridgway, E. C., Kliman, B., Maloof, F. 1976. Secretion of alpha subunit glycoprotein hormones by pituitary adenomas. *J. Clin. Endocrinol. Metabol.* 43:97-106
- Franks, S., Nabario, J. D. N. 1977. Prevalence and presentation of hyperprolactinaemia in patients with "functionless" pituitary tumors. *Lancet* 1:778-80
- McCormick, W. F., Halmi, N. S. 1971. Absence of chromophobe adenomas from a large series of pituitary tumors. *Arch. Pathol.* 92:231-38
- Guyda, H., Robert, F., Colle, E., Hardy, J. 1973. Histologic, ultrastructural, and hormonal characterization of a pituitary tumor secreting both hGH and prolactin. *J. Clin. Endocrinol. Metabol.* 36:531-47
- Ezrin, C., Kovacs, K. 1978. *Med. Clin. North Am.* 62:393-408
- Cunningham, G. R., Hucksins, C. 1977. An FSH and prolactin-secreting pituitary tumor: pituitary dynamics and testicular histology. *J. Clin. Endocrinol. Metab.* 44:248-53
- Horn, L., Erhardt, F., Fahlbusch, R., Pickardt, C. R., Werder, K. V., Schiba, P. C. 1976. Recurrent goiter, hyperthyroidism, galactorrhea and prolactin-

- producing pituitary tumor. *J. Clin. Endocrinol. Metab.* 43:137-43
19. Kleinberg, D. L., Noel, G. L., Frantz, A. G. 1977. Galactorrhea: a study of 235 cases, including 48 pituitary tumors. *N. Engl. J. Med.* 296:589-600
 20. Corenblum, B., Sirek, A. M. T., Horvath, E., Kovacs, K. L., Ezrin, C. 1976. Human mixed somatotrophic and lactotrophic pituitary adenomas. *J. Clin. Endocrinol. Metab.* 42:853-63
 21. MacLeod, R. M. 1976. Regulation of prolactin secretion. *Front. Neuroendocrinol.* 4:169-89
 22. Child, D. F., Nader, S., Mashiter, J., Kjeld, M., Banks, L., Fraser, T. R. 1975. Prolactin studies in "functionless" pituitary tumors. *Br. Med. J.* 1:604-6
 23. Malarkey, W. B. 1975. Nonpuerperal lactation and normal prolactin regulation. *J. Clin. Endocrinol. Metab.* 40:198-204
 24. Thorner, M. O., Besser, G. M., Jones, A., Dacie, J., Jones, A. E. 1975. Bromocriptine treatment of female infertility: report of 13 pregnancies. *Br. Med. J.* 4:694-97
 25. Buckman, M. T., Kaminsky, N., Conway, M., Peake, G. T. 1972. Utility of L-dopa and water loading in evaluation of hyperprolactinemia. *J. Clin. Endocrinol. Metab.* 36:911-19
 26. Adler, R. A., Noel, G. L., Wartofsky, L., Frantz, A. G. 1975. Failure of oral water loading and intravenous hypotonic saline to suppress plasma prolactin in man. *J. Clin. Endocrinol. Metab.* 41:383-89
 27. McIntosh, E. N. 1976. Treatment of women with the galactorrhea-amenorrhea syndrome with pyridoxine (vitamin B₆). *J. Clin. Endocrinol. Metab.* 42:1192-95
 28. Tolis, G., Laliberté, R., Guyda, H., Naftolin, F. 1977. Ineffectiveness of pyridoxine (B₆) to alter secretion of growth hormone and prolactin and absence of therapeutic effects on galactorrhea-amenorrhea syndromes. *J. Clin. Endocrinol. Metab.* 44:1197-99
 29. Goodenow, T. J., Malarkey, W. B. 1978. Ineffectiveness of pyridoxine in evaluation and treatment of the hyperprolactinemic amenorrhea-galactorrhea syndrome. *Am. J. Obstet. Gynecol.* In press
 30. Kleinberg, D. L., Noel, G. L., Frantz, A. G. 1971. Chlorpromazine stimulation and L-dopa suppression of plasma prolactin in man. *J. Clin. Endocrinol.* 33:873-76
 31. Lamberts, S. W. J., Birkenhäger, J. C., Kwa, H. G. 1976. Basal and TRH-stimulated prolactin in patients with pituitary tumors. *Clin. Endocrinol. Oxford* 5:709-11
 32. Malarkey, W. B., Johnson, J. C. 1976. Pituitary tumors and hyperprolactinemia. *Arch. Intern. Med.* 136:40-44
 33. Malarkey, W. B., Schroeder, L. L., Stevens, V. C., James, A. G., Lanese, R. R. 1977. Disordered nocturnal prolactin regulation in women with breast cancer. *Cancer Res.* 37:4650-54
 34. Krieger, D. T., Howanitz, P. F., Frantz, A. G. 1976. Absence of nocturnal elevation of plasma prolactin concentrations in Cushing's disease. *J. Clin. Endocrinol. Metab.* 42:260-72
 35. Schroeder, L. L., Johnson, J. C., Malarkey, W. B. 1976. Cerebrospinal fluid prolactin: a reflection of abnormal prolactin secretion in patients with pituitary tumors. *J. Clin. Endocrinol. Metab.* 43:1255-60
 36. Jordan, R. M., Kendall, J. W., Seaich, J. L., Allen, J. P., Paulsen, C. A., Kerber, C. W., Vanderlaan, W. P. 1976. Cerebrospinal fluid hormone concentration in the evaluation of pituitary tumors. *Ann. Intern. Med.* 85:49-55
 37. Assies, J., Schellekens, P. M., Touber, J. L. 1978. Prolactin in human cerebrospinal fluid. *J. Clin. Endocrinol. Metab.* 46:576-86
 38. Neelon, F. S., Goree, J. A., Lebovitz, H. E. 1973. The primary empty sella: clinical and radiographic characteristics and endocrine function. *Medicine* 52:73-92
 39. Vezina, J. L., Sutton, T. J. 1974. Prolactin-secreting pituitary micro-adenomas. *Am. J. Roentgenol.* 120:46-54
 40. Gomez, F., Reyes, F. I., Faiman, C. 1977. Nonpuerperal galactorrhea and hyperprolactinemia. *Am. J. Med.* 62:648-60
 41. Fine, S. A., Frohman, L. A. 1978. Loss of central nervous system component of dopaminergic inhibition of prolactin secretion in patients with prolactin-secreting pituitary tumors. *J. Clin. Invest.* 61:973-80
 42. Malarkey, W. B., Groshong, J. C. 1977. Defective dopaminergic regulation of prolactin secretion in a rat pituitary tumor cell line. *Nature* 266:640-41

STUDIES OF ATRIAL FLUTTER FOLLOWING OPEN HEART SURGERY¹

♦7321

*Albert L. Waldo, M.D., James L. Wells, Jr., M.D.,
Vance J. Plumb, M.D., Terry B. Cooper, M.D.,
and William A. H. MacLean, M.D.*

Department of Medicine and The Cardiovascular Research and Training
Center, University of Alabama in Birmingham, The School of Medicine,
Birmingham, Alabama 35294

INTRODUCTION

Patients frequently experience atrial flutter immediately after open heart surgery (1). At the University of Alabama Medical Center, all patients undergoing open heart surgery have a pair of temporary wire electrodes routinely placed on the atria at the time of operation for potential diagnostic and/or therapeutic use in the immediate postoperative period. The presence of a pair of fixed electrodes on the atria and the frequent occurrence of atrial flutter postoperatively have provided a unique opportunity to learn more about this long recognized (2) but poorly understood arrhythmia.

ENTRAINMENT AND INTERRUPTION OF CLASSICAL ATRIAL FLUTTER WITH RAPID ATRIAL PACING

As recently summarized (3), many studies used rapid atrial pacing utilizing a catheter electrode to treat atrial flutter. In most instances, atrial pacing at rates greatly in excess of the spontaneous atrial flutter rate produced atrial fibrillation, which, in turn, usually reverted spontaneously

¹Supported in part by USPHS NHLBI Program Project Grant HL11,310 and SCOR for Ischemic Heart Disease Grant 1P17HL17667 and by a Grant-in-Aid from the American Heart Association. Work was performed during Dr. Waldo's tenure as The Otto G. Storm Established Investigator of the American Heart Association.

to sinus rhythm. There were also several reports of failure to interrupt atrial flutter with rapid atrial pacing (4-7).

Critical Pacing Rate

As recently reported (8), systematic studies in patients who developed classical atrial flutter following open heart surgery demonstrated that to convert this rhythm to sinus rhythm with atrial pacing, the atria had to be paced at a critical rate for a critical duration of time. When the atria were paced from sites high in the right atrium at rates faster than that of the spontaneous atrial flutter but slower than the critical rate, the atrial rate was entrained to the faster pacing rate. This entrainment was transient, since with cessation of pacing or with slowing the pacing rate below the spontaneous atrial flutter rate, the atrial flutter returned promptly at its previous spontaneous rate. However, when the critical pacing rate was achieved, the atrial flutter was always interrupted (Figures 1-3).

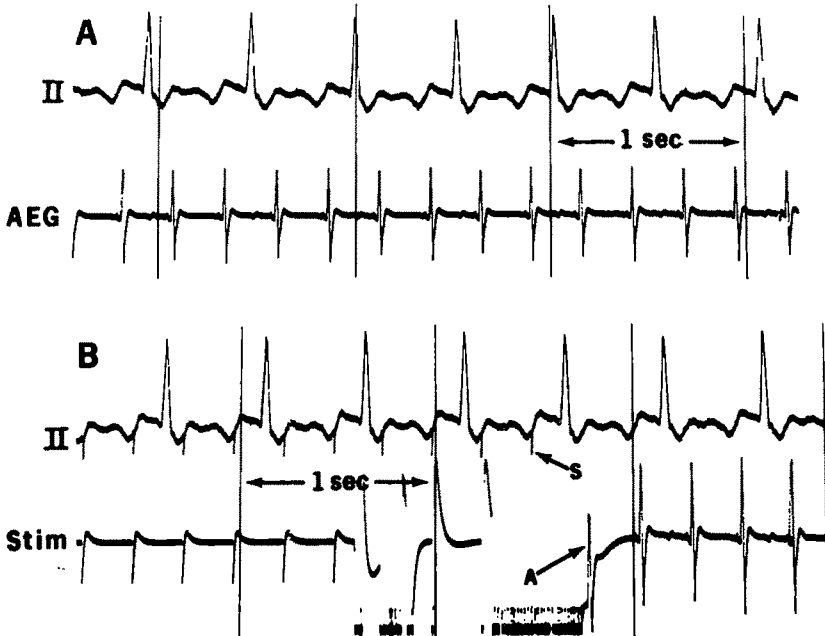


Figure 1 (A) ECG lead II recorded simultaneously with a bipolar atrial electrogram (AEG) from a patient who developed Type I (classical) atrial flutter following open heart surgery. The mean beat-to-beat atrial cycle length is 264 msec. (B) ECG lead II recorded simultaneously with the stimulus artifact (Stim) in the same patient at the end of 30 seconds of rapid atrial pacing from a high right atrial site at a cycle length of 254 msec. Note that during atrial pacing, although the atrial rate has increased to the pacing rate, the morphology of the atrial complexes in the ECG is unchanged. S = stimulus artifact, A = atrial electrogram. Times lines at 1-sec intervals. Modified after Waldo et al (8).



Figure 2 ECG lead II recorded simultaneously with the stimulus artifact (Stim) in the same patient as in Figure 1 recorded (A) at the end of 30 seconds of atrial pacing at a cycle length of 242 msec and (B) at the end of 30 seconds of atrial pacing at a cycle length of 232 msec. Note that although the atrial rate has increased to the pacing rate in both A and B, the morphology of the spontaneous flutter waves in the ECG is very little different from that during atrial pacing, except for some minimal change in flutter wave morphology at the cycle length of 232 msec. S = stimulus artifact, A = atrial electrogram. Modified after Waldo et al (8).

Critical Polarity and Morphology of Atrial Complexes in the ECG

The hallmark of the successful interruption of atrial flutter with rapid atrial pacing from sites high in the right atrium was the change in polarity of the atrial complexes in ECG leads II, III, and aVF from their previously negative or predominantly negative polarity to a positive polarity (3, 8). This is illustrated in Figures 1–3. It is remarkable that during the period of transient entrainment, i.e. during the period when the atrial rate increased to the pacing rate but the atrial flutter was not interrupted, the morphology and polarity of the atrial complexes in ECG leads II, III, and aVF were virtually identical to, if not in fact identical to, those during the spontaneous rhythm. This occurred despite the fact that the atria were being paced from sites that normally produce positive atrial complexes in these leads (9, 10). Why this is so is not established, but clearly, whatever

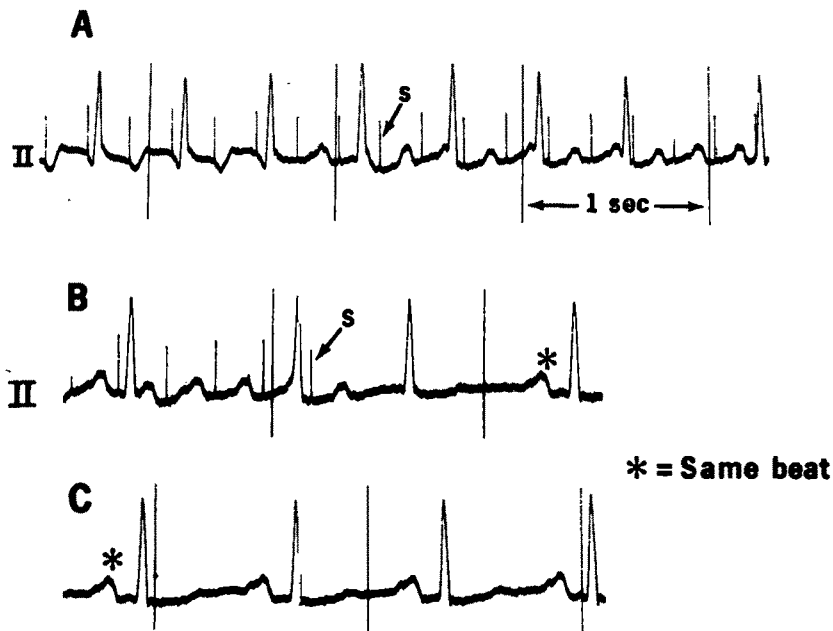


Figure 3 (A) ECG lead II recorded in the same patient as in the preceding two figures during atrial pacing at a cycle length of 224 msec. Note that with the seventh atrial beat in this tracing, and after 22 sec of atrial pacing, the atrial complexes suddenly became positive. (B) ECG lead II recorded at the termination of atrial pacing in the same patient. Note that with abrupt termination of pacing, sinus rhythm recurs. (C) The first beat in this panel is identical with the last beat in B. Modified after Waldo et al (8).

generates the atrial flutter is capable of increasing its rate to the pacing rate in such a fashion that the sequence of atrial activation apparently is not significantly changed.

In some cases during the transient entrainment of atrial flutter, as the atrial pacing rate approached the critical rate for interruption of atrial flutter, the atrial complexes in leads II, III, and aVF appeared to have a morphology intermediate between that characteristic of the spontaneous atrial flutter and that characteristic of pacing from a high right atrial site, i.e. there appeared to be a fusion morphology (8). In some cases, when pacing the atria at the critical rate to achieve interruption of the atrial flutter, this apparent fusion morphology appeared transiently prior to the appearance of frankly positive atrial complexes in these same ECG leads (Figure 4). Following the appearance of this fusion morphology, if the rapid atrial pacing was stopped or slowed to a rate slower than the spontaneous atrial flutter rate, the spontaneous atrial flutter reappeared (8). Finally, for some patients, while rapid atrial pacing initially entrained the atrial flutter,

Pace Atria 320 beats/min

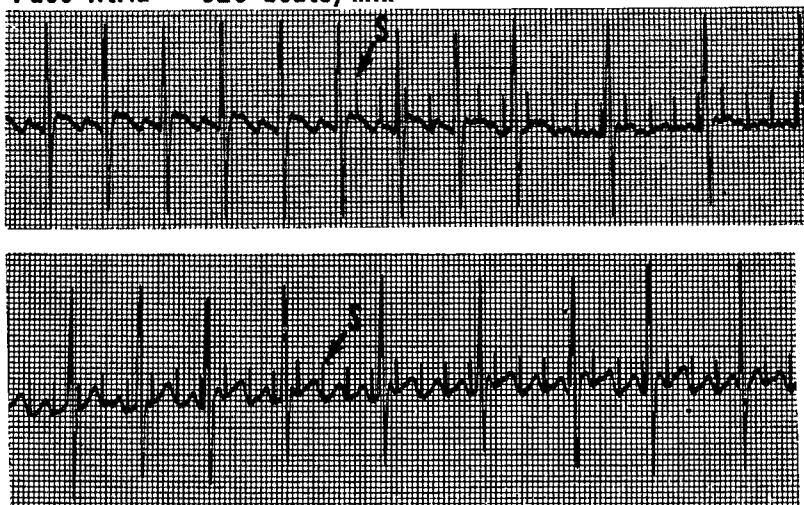


Figure 4 ECG lead II recorded from a patient who developed atrial flutter at a rate of about 280 beats/min following open heart surgery. In the top tracing, atrial pacing at a rate of 320 beats/min was initiated. With the onset of pacing, the atrial complexes in the ECG that had been predominantly negative became predominantly positive and notched. The bottom trace was recorded several seconds after the top trace. Note that now, although the atrial pacing rate remains at 320 beats/min, the atrial complexes are completely positive. When the atrial pacing rate was subsequently slowed to 120 beats/min, the atrial rate remained controlled by the pacing rate and when the pacing was then abruptly terminated, a spontaneous sinus rhythm of 100 beats/min obtained. Previously, when pacing was slowed after the atrial complexes in the top tracing appeared but before the atrial complexes in the bottom tracing appeared, the atrial flutter was not interrupted. S = stimulus artifact. From Cooper et al (3).

still more rapid atrial pacing continued to entrain the atrial flutter until continued increase in the pacing rate produced atrial fibrillation (3, 8). If it was not the previously predominant rhythm, the atrial fibrillation produced usually was transient, reverting spontaneously to sinus rhythm, or in a small percentage of cases, reverting back to atrial flutter.

Insights into the Nature of Classical Atrial Flutter

These data suggest that whatever the mechanism of classical atrial flutter, be it reentrant, automatic (? triggered automaticity), or a yet unidentified mechanism, atrial flutter can be sustained in one part of the atria even when other parts are being depolarized from a site high in the right atrium. This makes less tenable the old hypothesis that atrial flutter is a rhythm generated by reentry around the great veins (11). Additionally, based on our own findings (9, 10) and those of others (12–17) that negative atrial complexes

can be generated by pacing inferior sites in both atria, these data suggest that classical atrial flutter is generated at a focus low in either the right or left atrium. It should be emphasized that use of the term focus does not necessarily imply an automatic mechanism. Although some form of an automatic mechanism is possible, it is also possible that the responsible focus consists of a small circus movement (reentry) around one of the vessels that enters the atria inferiorly (such as the inferior vena cava, one or more of the pulmonary veins, or the coronary sinus) or around a small circuit of atrial muscle (18).

Critical Duration of Pacing

The duration of pacing is also important in the successful interruption of atrial flutter (Figures 3 and 4) (3, 8). Thus, even when pacing from sites high in the right atrium at the appropriate rate for interruption of the atrial flutter, the mean duration of pacing required was 10 sec (range 2–22 sec). This has been repeatedly confirmed by our subsequent observations, which include one instance that required 39 sec of pacing at a constant rate (A. L. Waldo, unpublished observation).

The reason for a necessary minimal duration of sustained atrial pacing is unclear. It may be related to the location of the atrial pacing wires in our patients (3, 4, 8). For instance, if atrial flutter is generated at a focus low in the left or right atrium, the distance of the pacing site from the flutter focus may influence the necessary duration of pacing at the critical rate. Additionally, the fact that conduction in the atria at rates in the range of atrial flutter is much slower than at normal rates (19) is very likely another important factor. Similarly, the critical duration of atrial pacing may relate in part to a beat-to-beat increase in the extent of atrial tissues activated by a wavefront coming from the pacing site, taking a relatively long time until it can overcome the atrial activation that results from a wavefront from the flutter focus.

Critical Pacing Threshold

Failure to interrupt atrial flutter with rapid atrial pacing techniques may result from pacing at an inadequate rate or for an inadequate period of time. It has also undoubtedly been caused by failure to capture the atria. Barring technical problems with equipment (e.g. dead pacemaker battery, broken electrode, or the like), one of the most common problems in achieving atrial capture is failure of the electrode to maintain contact with atrial tissue when pacing at rapid rates using the catheter electrode technique (3). The use of wire electrodes temporarily fixed to the atria in post open heart surgical patients obviated this problem and also permitted the demonstration that when pacing the atria at very rapid rates, i.e. rates in the range required to interrupt atrial flutter, the pacing threshold required for atrial capture is

very high (19). It is usually between 10–20 mA and on occasion may be even higher. The highest documented threshold required to capture the atria in order to interrupt atrial flutter is 49 mA (A. L. Waldo, unpublished observation). Thus, another possible reason for reported failure to interrupt classical atrial flutter with rapid atrial pacing is inadequate strength of the pacing stimulus.

TYPES OF ATRIAL FLUTTER

We identified two types of atrial flutter in patients after open heart surgery: Type I (classical) and Type II (3, 20). This classification is independent of the morphology or polarity of the flutter waves in the ECG.

Characteristics of Type I and Type II Atrial Flutter

SIMILARITIES Both types of atrial flutter are characterized by a remarkably constant beat-to-beat morphology, polarity, amplitude, and cycle length of the recorded bipolar atrial electrogram (Figures 1 and 5) (20). The variability in beat-to-beat cycle length is small; the overall mean for both types is 5.0 msec with a range of 1.2–16.5 msec in the patients studied. A beat-to-beat electrical alternans was recorded in some patients in both groups (20). In some instances, the electrical alternans was associated with two constant populations of cycle lengths. Within these two populations, the beat-to-beat cycle lengths varied very little, and were virtually identical to the mean beat-to-beat variability for all patients. Thus, the few instances of apparent relatively large beat-to-beat differences in cycle length (e.g. 16.5 msec) were misleading.

DIFFERENCES Type I atrial flutter could always be influenced by rapid atrial pacing from sites high in the right atrium, whereas Type II could not (3, 20). This characteristic of Type II atrial flutter provides yet another possible reason for reported failures to interrupt atrial flutter with rapid atrial pacing techniques. It may be that Type II atrial flutter is not in-

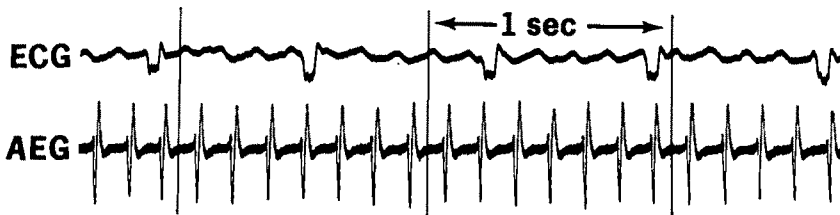


Figure 5 ECG lead III recorded simultaneously with a bipolar atrial electrogram (AEG) demonstrating Type II atrial flutter at an atrial rate of 420 beats/min. Modified after Wells et al (20).

fluenced by rapid atrial pacing because the pacing site is so far from the flutter focus that the factors of atrial conduction time and atrial refractoriness preclude the interruption of Type II flutter in this fashion. Thus, if the only difference between Types I and II atrial flutter was the response to rapid atrial pacing from sites high in the right atrium, the differentiation of atrial flutter into two types would be less tenable.

Two other observations support separation of the two types of atrial flutter. First, Type II atrial flutter was observed to convert spontaneously to Type I in a stepwise fashion (20). Second, on several occasions Type I atrial flutter was converted to Type II by rapid atrial pacing (20).

Based on the above differences, Type I atrial flutter was characterized by a range of rates from 240–340 beats/min and Type II atrial flutter from 340–433 beats/min (3, 20). Undoubtedly, these ranges of rates are not precise, and there is very likely some overlap in their ranges. Also, the lower range for Type I probably includes rates somewhat lower than 240 beats/min and the upper range of rate for Type II probably includes rates somewhat higher than 433 beats/min.

Differences Between Atrial Flutter and Atrial Fibrillation

The differentiation between atrial flutter and atrial fibrillation might seem obvious and well described. However, our studies of atrial fibrillation following open heart surgery (21) demonstrated that two of the four types of this arrhythmia that are relatively nonchaotic (Type I atrial fibrillation in which the recorded atrial electrogram demonstrates distinct atrial complexes with an isoelectric baseline, and Type II atrial fibrillation in which the electrogram demonstrates distinct atrial complexes but with perturbations of the baseline between atrial complexes) may resemble atrial flutter (Figure 6). However, the hallmark of atrial fibrillation, the marked change



Figure 6 ECG lead III recorded simultaneously with a bipolar atrial electrogram demonstrating Type I atrial fibrillation. Note the discrete atrial complexes in the bipolar atrial electrogram recording. Further note that these complexes are of variable morphology, polarity, amplitude, and cycle length. From Wells et al (21).

in beat-to-beat morphology, polarity, amplitude, and cycle length of the recorded bipolar atrial electrogram, clearly and readily distinguishes it from atrial flutter (20, 21).

CONCLUSIONS

On the basis of studies of postoperative, open heart surgery patients, two types of atrial flutter, Types I and II, have been identified. They resemble each other in that they are characterized by a regular beat-to-beat morphology, polarity, amplitude, and cycle length of the recorded bipolar atrial electrogram. They differ in terms of the range of rates (Type II is faster than Type I atrial flutter) and in the response to rapid atrial pacing from high atrial sites (Type I is always influenced and Type II atrial flutter is uninfluenced by such pacing). To interrupt Type I atrial flutter with rapid atrial pacing from a high right atrial site requires critical atrial pacing rate, duration of pacing, and stimulus strength. Analysis of the morphology and polarity of atrial complexes in the ECG during rapid atrial pacing of Type I atrial flutter demonstrated transient entrainment of the flutter prior to its interruption, and permitted inferences about the mechanism of the flutter and the location of the flutter focus.

Literature Cited

1. Waldo, A. L., MacLean, W. A. H., Cooper, T. B., Kouchoukos, N. T., Karp, R. B. 1978. The use of temporarily placed epicardial atrial wire electrodes for the diagnosis and treatment of cardiac arrhythmias following open heart surgery. *J. Thorac. Cardiovasc. Surg.* 76:500-5
2. Jolly, W. A., Ritchie, W. T. 1910-1911. Auricular flutter and fibrillation. *Heart* 2:177-221
3. Cooper, T. B., MacLean, W. A. H., Waldo, A. L. 1978. Overdrive pacing for supraventricular tachycardia: A review of theoretical implications and therapeutic techniques. *PACE* 1:196-221
4. Wellens, H. J. J. 1971. *Electrical Stimulation of the Heart in the Study and Treatment of Tachycardias*. pp. 27-39. Baltimore: Univ. Park Press. 142 pp.
5. Rosen, K. M., Sinno, M. Z., Gunnar, R. M., Rahimtoola, S. H. 1972. Failure of rapid atrial pacing in the conversion of atrial flutter. *Am. J. Cardiol.* 29:524-28
6. Zipes, D. P. 1971. The contribution of artificial pacemaking to understanding the pathogenesis of arrhythmias. *Am. J. Cardiol.* 28:211-22
7. Haft, J. I. 1974. Treatment of arrhythmias by intracardiac electrical stimulation. *Prog. Cardiovasc. Dis.* 16:539-68
8. Waldo, A. L., MacLean, W. A. H., Karp, R. B., Kouchoukos, N. T., James, T. N. 1977. Entrainment and interruption of atrial flutter with rapid atrial pacing. Studies in man following open heart surgery. *Circulation* 56:737-45
9. Waldo, A. L., Vitikainen, K. J., Kaiser, G. A., Malm, J. R., Hoffman, B. F. 1970. The P wave and P-R interval. Effects of the site of origin of atrial depolarization. *Circulation* 42:653-71
10. MacLean, W. A. H., Karp, R. B., Kouchoukos, N. T., James, T. N., Waldo, A. L. 1975. P waves during ectopic atrial rhythms in man. A study utilizing atrial pacing with fixed electrodes. *Circulation* 52:426-34
11. Lewis, T., Drury, A. N., Iliescu, C. C. 1921. A demonstration of circus movement in clinical flutter of the auricles. *Heart* 8:341-59
12. Rosen, K. M., Lau, S. H., Damato, A. N. 1969. Simulation of atrial flutter by rapid coronary sinus pacing. *Am. Heart J.* 78:635-42

13. Frankl, W. S., Soloff, L. A. 1968. Left atrial rhythm. Analysis by interatrial electrocardiogram and vectorcardiogram. *Am. J. Cardiol.* 22:645-56
14. Gonzalez Videla, J. 1968. El diagnostico del origen del estimulo por los caracteres de la deflexion auricular. Evaluacion de los signos diagnosticos de los ritmos auriculares izquierdos. *Rev. Argent. Cardiol.* 35:105-21
15. Harris, B. C., Shaver, J. A., Gray, S., Kroetz, F. W., Leonard, J. J. 1968. Left atrial rhythm: Experimental production in man. *Circulation* 37:1000-14
16. Lancaster, J. F., Leonard, J. J., Leon, D. F., Kroetz, F. W., Shaver, J. A. 1965. The experimental production of coronary sinus rhythm in man. *Am. Heart J.* 70:89-93
17. Lau, S. H., Cohen, S. I., Stein, E., Haft, J. I., Rosen, K. M., Damato, A. N. 1971. P waves and P loops in coronary sinus in left atrial rhythms. *Am. Heart J.* 79:201-14
18. Allessie, M. A., Bonke, F. I., Schopman, F. J. G. 1977. Circus movement in rabbit atrial muscle as a mechanism of tachycardia. III. The "leading circle" concept: A new model of circus movement in cardiac tissue without the involvement of anatomical obstacle. *Circ. Res.* 41:9-18
19. Plumb, V. J., Karp, R. B., Cooper, T. B., MacLean, W. A. H., James, T. N., Waldo, A. L. 1978. Atrial conduction and excitability during rapid atrial pacing in man. *Clin. Res.* 26:46A (Abstr.)
20. Wells, J. L. Jr., MacLean, W. A. H., James, T. N., Waldo, A. L. 1978. Characterization of atrial flutter. Studies in man following open heart surgery using fixed atrial electrodes. *Circulation*. Submitted
21. Wells, J. L. Jr., Karp, R. B., Kouchoukos, N. T., MacLean, W. A. H., James, T. N., Waldo, A. L. 1978. Characterization of atrial fibrillation in man. Studies following open heart surgery. *PACE* 1:426-38

CONTROL OF THE IMMUNE SYSTEM BY INHIBITOR AND INDUCER T LYMPHOCYTES

◆7322

H. Cantor, M.D.

Department of Medicine, Harvard Medical School,
Boston, Massachusetts 02115

The field of immunology has its historical roots in clinical medicine. It grew directly out of modern medicine's greatest success: diagnosis and prevention of bacterial diseases, which began at the turn of the century. Since then the field has expanded enormously. Analysis of the immunologic system is now recognized as a distinct field of biology. Studies of the biochemical regulation of immunologic cells promise to be the key approach to understanding how genes work in mammalian cells. Nonetheless, the historical connection between immunology and clinical medicine has been maintained as well, perhaps more vigorously than in any other field of biology. There is a constant interchange among scientists doing "basic" immunologic research and those engaged in "applied" research who study the disease processes that ensue when the immune system malfunctions. Over the past ten years, an explosion of information from basic experimental studies has resulted in an entirely new view of the immune system that promises to have an important impact upon applied or clinical immunology. The following is a summary of some of these developments.

The important cells in the immune system are lymphocytes and macrophages. Macrophages line the tissues of the body, ingest foreign material, and present it to lymphocytes. Lymphocytes circulate freely through the blood and lymphatic vessels of the body and bear clonally distributed receptors for antigen. Lymphocytes are directly responsible for all specific immune responses. They are divisible into two sorts: (a) the T lymphocytes

(or T cells), so-called because their maturation requires processing in the thymus; and (b) the B lymphocytes (or B cells), which in mammals are probably generated mainly in bone marrow.

It should be emphasized that although the field of immunology began at the turn of the century, identification of the lymphocyte as the immunological cell was clearly established only 15–20 years ago. This means that cellular immunology is an extraordinarily young discipline, and a coherent view of the cellular basis of the immune response is only just beginning to emerge. Until recently, immunologists generally held that the immune system was comprised of clones of lymphocytes that, when activated by antigen, produced antibody or initiated a cell-mediated response (such as an inflammatory response). According to this view, the duration and strength of a response depended only on the number of lymphocyte clones in the host that carried receptors for a particular antigen, and the absence of immunity to “self” reflected deletion at birth of all lymphocytes bearing receptors for self-antigens.

Over the past decade, evidence refuting this view of the immune system has increased. More accurately described, the immune system is composed of many sets of regulatory lymphocytes that respond mainly to signals generated from within the system itself, and for the most part these interactions inhibit both antibody and cellular immune responses; these interactions also serve to prevent B lymphocytes from producing antibody against host antigens. In other words, the absence of autoimmune reactions may be, in part, due to continuous and active suppression rather than the absence of self-reactive cells within the system.

One approach supporting this view of the immune system involves the dissection and definition of T cells and the various immune functions they perform. Despite their uniform morphology, T cells are by no means a homogeneous population; they comprise subclasses or sets of lymphocytes with different and even seemingly opposing functions. Thus, one property of T cells, called helper function, is to assist B cells to make antibody. A second function of T cells was also suspected from investigations of immunologic tolerance or unresponsiveness. These studies indicated that the adoptive transfer of T lymphocytes from an animal unresponsive to a given antigen to a normal animal could render the recipient specifically unresponsive. This property of T cells, termed “suppressor” function, has been subsequently observed in a large number of immunologic experiments. It is likely that suppressor function is a necessary homeostatic control mechanism that keeps the immune system “in trim” and prevents untoward autoimmune reactions. Another property of T cells is the generation of cells capable of damaging or destroying those cells recognized as antigenically foreign after, for example, infection by a virus. This is associated with the

cytotoxic or killer function of T cells. Yet another function of T cells involves inducement or activation of other cells to participate in inflammatory responses.

A crucial point arises: Are all these functions invested in a single set of T cells differentiated in the thymus, and are these diverse responses governed entirely by extraneous conditions such as a mode or type of antigen stimulation? Alternatively, are these immunologic functions invested in distinct sets of T cells programmed to respond in different ways during their differentiative history? According to this latter idea, thymus-dependent differentiation may give rise to a number of separate sublines of mature T cells, each genetically programmed to mediate one or another T-cell response.

In the mouse, this question has resolved itself into the practical problem of finding out whether it is possible to subdivide the T-cell population into different sets that, when confronted with antigens, are able to make only one or another of the possible T-cell responses. Presently, the most effective technique for identifying and separating subpopulations of peripheral T cells comes from studies of the cell surface components, which are expressed on cells undergoing thymus-dependent differentiation. This classification is based upon the use of alloantisera to define cell surface differentiation components expressed on T cells. Since these components have not been detected on cells of other tissues such as brain, kidney, liver, or epidermal cells, they are evidently specified by genes expressed exclusively during T-cell differentiation. These are called the Ly systems. The Ly1 component is coded for by a gene on chromosome 19, and the Ly2,3 components are both coded for by genes on chromosome 6. These last two are tentatively treated together because the two genes are tightly linked, and these two systems have not yet exhibited any differences other than the fact that they are genetically coded for by distinguishable loci.

In general, the approach involves a cytotoxicity assay similar to the complement-dependent hemolytic test used to identify markers on red cells. As with hemolysis by antierythrocyte antibody and complement, lymphocytes exposed to, say, anti-Ly1 sera in the presence of complement are lysed. This lysis can be monitored by the use of trypan blue, which stains lysed but not living cells, or by the release of a radioactive label from the lysed cells. More recently, these antisera to Ly1 or Ly2 components have been used to "positively" select cells bearing these components: Columns containing beads coated with anti-Ly1 or anti-Ly2 selectively retain lymphocytes expressing the relevant Ly surface component.

Analysis of this sort revealed that the peripheral T-cell pool contains at least three separate T-cell sets. We refer to them as the Ly123 set, the Ly1 set, and the Ly23 set. They compose respectively 50%, 30%, and approxi-

mately 10% of the peripheral T-cell pool. Thus, according to the criterion of selective expression of gene products on the cell surface, the T-cell pool is divisible into three groups of cells, each following a different set of genetic instructions. The question then becomes whether these individual differentiative programs include information that decides what the function of each T-cell set should be. Evidence to date indicates that cells of the Ly1 set are genetically programmed to help or amplify activity of other cells after stimulation by antigen. Ly1 cells are most aptly called "inducer" cells since they will induce or activate other cell sets to fulfill their respective genetic programs: Ly1 cells induce B cells to secrete antibody; they induce macrophages and monocytes to participate in delayed-type-hypersensitivity responses; they can, under appropriate circumstances, induce precursors of killer cells to differentiate to killer-effector cells; most recently, and perhaps most importantly, Ly1 cells were also found to induce a set of resting, nonimmune T cells to generate potent "feedback" suppressive activity. Analysis of isolated Ly1 inducer cells from nonimmune donors indicates that these cells are already programmed for helper/inducer function *before* overt immunization with antigen.

By contrast, cells of the Ly23 set are specially equipped both to develop alloreactive cytotoxic activity and to suppress humoral and cell-mediated immune responses. Whether cytotoxicity and suppression are two manifestations of one genetic program or whether they represent the phenotype of two separate genetic programs is not yet established.

It is of particular interest that both sets of T cells see antigen in the context of the host's own major histocompatibility (MHC) gene products. In other words, these T cells screen the host for "foreignness" as judged by alterations of their own MHC antigens as, for example, when these MHC products have been "modified" by association with a virus. Ly1 cells are selectively activated and bind to the so-called I-region products of the MHC, which correspond to HLA-D products in man. I-region products may be recognized as foreign because of either polymorphic variation (alloantigens) or modification by foreign antigens such as viruses. By contrast, Ly23 cells bind to and react against H2/KD gene products of the MHC which roughly correspond to HLA-A and B antigens in man.

Proof that these two cell sets, which are marked by different surface Ly phenotypes and different functional potentials, in fact represent two branches of thymus-directed differentiation comes from experiments in which isolated Ly1 cells and Ly23 cells are used to repopulate the lymphoid tissues of mice depleted of their T-cell system. These recipients lack detectable numbers of T cells and are called B mice. Recipients of Ly1 cells are called B-Ly1 mice and recipients of Ly23 cells are called B-Ly23 mice. For as long as we have observed them, B-Ly1 mice have been equipped for helper function but not killer function, and B-Ly23 mice have been

equipped for primary killer function but not for helper function. These findings show that Ly1 and Ly23 cells have already exercised differentiative options that prevent them from giving rise to one another. In other words, these two T-cell sets belong to different lines of differentiation and are not sequential stages of a single progression.

Until recently cells of the Ly123 set have been the least well defined of the various T-cell sets. Most likely at least some Ly123 cells represent a store of receptor-positive intermediary cells that regulate the supply and function of more mature Ly1 and Ly23 cells. This is supported in part by experiments showing that (a) after stimulation with virus-infected syngenic cells, some Ly123 cells give rise to Ly23 progeny; and (b) purified populations of Ly123 cells can give rise to Ly1 cells after polyclonal activation by concanavalin A. That at least a portion of Ly123 cells represent a precursor pool is also consistent with earlier observations that cells of the Ly123 subclass are detectable in the spleens of mice within the first week of life, while neither Ly1 nor Ly23 cells reach maximal numbers until adult life (8–12 weeks of age).

More recently, it has been demonstrated that antigen-stimulated Ly1 cells, or supernatants of activated Ly1 cells, in addition to inducing B cells to secrete antibody, can induce or activate resting Ly123 T cells to develop profound feedback-suppressive activity. The term "feedback suppression" is appropriate here since the degree of suppressive activity exerted by a fixed number of nonimmune Ly123 cells increases in direct proportion to the numbers of antigen-activated Ly1 cells in the system. This Ly1:Ly123 interaction has also been shown to influence the immune response in vivo in mice and may well represent the central cell interaction that governs the duration and intensity of immune reactions: After stimulation of the immune system by foreign materials, activated antigen-specific Ly1 cells induce B cells to form antibody and also induce resting Ly123 cells to inhibit T-helper cell activity. Reduction in T-helper activity is accompanied by decreased induction of B cells as well as progressively decreasing induction of resting Ly123 cells; the net result is progressive decrease in both antibody formation and suppressor cell induction. These findings also indicate that, like the formation of antibody, the generation of immunologic suppression after stimulation by antigen is not an autonomous function; both require induction by Ly1 cells.

These negative feedback circuits suggest also that the response to any given antigen may reflect in part the amount of feedback inhibition generated after exposure to that antigen. The response to many antigens is regulated by MHC-linked genes: Each inbred strain of mouse will not respond to a certain proscribed list of antigens. In some cases, this lack of response reflects exaggerated induction of feedback-inhibitory cells that mask delivery of the T-helper signal to the B cell.

The ability of Ly1 inducer cells to activate various effector cell systems on the one hand, and suppressive systems on the other, is essential to the regulation of both the intensity and type of the immune response. It therefore has become critically important to know whether Ly1 cells that induce suppressive activity represent a specialized fraction of Ly1 cells. A direct approach to this question comes from the finding that some Ly1 cells also express a newly defined antigen called Qa1. This antigen, or antigen system, is coded for by genes that are closely linked to MHC genes of the mouse. Studies of Ly1:Qa1⁺ cells show that these cells are responsible for induction of feedback inhibition and that Ly1:Qa1⁻ cells are not. In addition, these studies show that signals from both Ly1:Qa1⁺ and Ly1:Qa1⁻ cells are required for optimal formation of antibody by B cells. Thus, the ability of antigenic determinants to induce a detectable antibody response may depend largely on the ratio of Ly1:Qa1⁺ and Ly1:Qa1⁻ T-cell clones that bear receptors for that antigen. Recent work also suggests that the ability of Ly1:Qa1⁺ inducer cells to elicit strong suppressive responses is likely to be particularly important in governing the duration and intensity of inflammatory reactions such as delayed-type hypersensitivity and IgE-mediated hypersensitivity. Analysis of the cell-free products of homogeneous populations of Ly1:Qa1⁺ cells is currently of intense interest since these materials may well prove useful in strategies designed to selectively suppress hypersensitivity or antibody responses to defined antigens.

This analysis indicates that perturbation of the immune system by antigen results in stimulation of two distinct inducer T-cell sets; both deliver signals that activate B cells to produce antibody, but only one activates the T-suppressive system, and it is the strength of this latter interaction that ultimately sets the level and the duration of an immune response.

In sum, these experiments established that (a) the genetic program for a single differentiated set of cells, in this case immunologic cells, combines information coding for a surface antigenic profile associated with particular physiologic functions; (b) the majority of T cells are not effector cells poised to respond to foreign antigen, but regulatory cells that respond mainly to signals or messages generated from within the T-cell system itself; and (c) detectable immune responses reflect perturbations of these signals after the Ly1 inducer system is stimulated by "antigen." The net effect of these T-T interactions after perturbation by antigen is to restore the homeostatic balance of the system, usually at a new level reflecting differentiation of antigen-specific T- and B-cell clones belonging to the sets described above.

We have just begun to delineate the circuits involved in the regulatory system summarized above. The picture beginning to emerge is that Ly1 cells act as sentinel cells that screen the surfaces of other cells, particularly macrophages, for foreign material associated with MHC molecules (Figure 1). When activated, these sentinel cells can induce a variety of effector cells

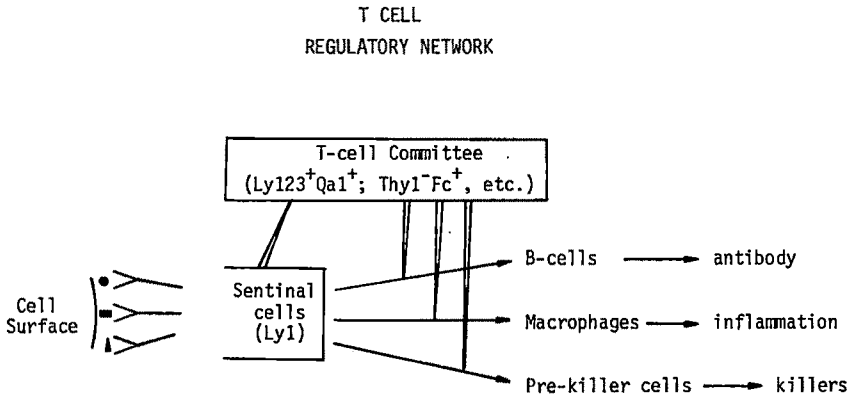


Figure 1 T-cell regulatory network.

(e.g. B cells that make antibody, or macrophages and monocytes that participate in inflammatory responses) to produce a specific immune response. In addition, they activate a "committee" of resting T cells that are probably relatively immature.¹ This committee of cells emits inhibitory signals. The intensity of inhibition depends mainly on the genetic background of the host, the nature of the antigenic stimulus, and the intensity of the inducing signal emitted by the sentinel cells. The observed immune response depends upon the relative potency and timing of feedback-suppressive inhibitory signals.

What happens when this system goes wrong? There are, so far, two examples of disorders of this immunoregulatory circuit. NZB mice spontaneously develop an autoimmune disorder characterized by the production of a variety of autoantibodies and a clinical syndrome resembling human systemic lupus erythematosus. The major T-cell deficit of NZB mice is the absence or malfunction of an Ly123 T-cell set responsible for feedback inhibition.

A second example comes from experiments in which Ly123 regulatory cells were deliberately eliminated from the host: Mice depleted of all T cells were repopulated with either Ly1 inducer cells or all T-cell sets (including Ly123 regulatory cells). Within the first two weeks after repopulation, sera from the former but not from the latter mice contained autoantibodies against erythrocytes and thymocytes. Thus, elimination of the regulatory T cells that participate in feedback inhibition results directly in the formation of autoantibodies.

¹The term "committee" is used here because it is virtually certain that Ly123 cells are themselves a heterogeneous set, and perhaps not the sole members of this system. Moreover, the response of a committee is generally suppressive.

It should be emphasized that a number of critical questions remain to be answered. Are feedback-inhibitory interactions among T-cell sets responsible, in part, for self-tolerance beginning at birth? What is the molecular basis underlying communication among these T-cell sets? And, finally, how can we use this information to improve our understanding and treatment of autoimmune disorders in man?

Recent studies using antisera specific for subsets of human T lymphocytes revealed that human T cells are separable into sets of functionally distinct populations that so far appear analogous to those identified in mice. One set exerts suppressive activity, as judged by the ability of these cells to inhibit the *in vitro* production of immunoglobulin by human B cells. A second set of cells helps human B cells to secrete immunoglobulin. Recently, it was demonstrated that sera from children suffering from rheumatoid arthritis contain antibodies that selectively react with suppressor T cells found in normal individuals. Moreover, patients with active disease, but not those in remission, have high titers of antippressive cell antibody and lack T cells that suppress the *in vitro* production of immunoglobulin by B cells. Whether removal of suppressor T cells by this autoantibody plays a key role in the pathogenesis of this disease remains to be established. Selective dysfunction of T-suppressor activity was also reported in studies of patients with systemic lupus erythematosus, and there is reason to suspect that other major diseases such as multiple sclerosis, certain types of diabetes mellitus, and perhaps a subset of essential hypertension may reflect, in part, disorders of the immunoregulatory system.

No one can say with certainty whether these new insights into the workings of the immune system will have an important impact upon treatment of autoimmune disorders in man. Dissection of regulatory T-cell circuits is currently the subject of intense experimentation and debate. And perhaps this is the most hopeful and encouraging aspect of this new phase of immunobiology.

Selected References

- Alter, B. J., Bach, F. H. 1974. Role of H-2 lymphocyte defined and serologically defined components in the generation of cytotoxic lymphocytes. *J. Exp. Med.* 140:1410
- Bach, F. H., Bach, M. L., Soudel, P. M. 1976. Differential function of major histocompatibility complex antigens and T lymphocyte activation. *Nature* 259:273
- Boyse, E. A., Old, L. J. 1969. Some aspects of normal and abnormal cell surface genetics. *Ann. Rev. Genet.* 3:269
- Broder, S., Humphrey, R., Durm, M., Blackman, M., Meade, B., Goldman, C., Strober, W., Waldmann, T. A. 1975. Impaired synthesis of polyclonal immunoglobulins by circulating lymphocytes from patients with multiple myeloma. *N. Engl. J. Med.* 293:887
- Chess, L., Schlossman, S. F. 1977. *Adv. Immunol.* pp. 125-213
- Cantor, H., Boyse, E. A. 1977. Regulation of cellular and humoral immunity by T-cell subclasses. In *Proc. 41st Cold Spring Harbor Symp. Origin of Lymphocyte Diversity*, Vol. XLI, p. 23. Cold Spring Harbor Lab., NY

- Cantor, H., McVay-Boudreau, L., Hugenberger, J., Naidorf, K., Shen, F. W., Gershon, R. K. 1978. Immunoregulatory circuits among T-cell sets. II. Physiologic role of feedback inhibition in vivo: Absence in NZB mice. *J. Exp. Med.* 147:1116
- Cantor, H., Weissman, I. 1976. Development and function of subpopulations of thymocytes and T cells. *Prog. Allergy* 20:1
- Eardley, D. D., Hugenberger, J., McVay-Boudreau, L., Shen, F. W., Gershon, R. K., Cantor, H. 1978. Immunoregulatory circuits among T-cell sets. I. T-helper cells induce other T-cell sets to exert feedback inhibition. *J. Exp. Med.* 147:1106
- Engleman, E. G., McMichael, A. J., Batey, M. E., McDevitt, H. O. 1978. A suppressor T cell of the mixed lymphocyte reaction in man specific for the stimulating alloantigen. *J. Exp. Med.* 147:137
- Evans, R. L., Breard, J. M., Lazarus, H., Schlossman, S. F., Chess, L. 1977. Detection, isolation and functional characterization of two human T cell subclasses bearing unique differentiation antigens. *J. Exp. Med.* 145:2221
- Evans, R. L., Lazarus, H., Penta, A. C., Schlossman, S. F. 1978. Two functionally distinct subpopulations of human T cells that collaborate in the generation of cytotoxic cells responsible for cell-mediated lympholysis. *J. Immunol.* 120:1423
- Good, R. A. 1973. Immunodeficiency in developmental perspective. *The Harvey Lecture Series* 67, p. 1. New York: Academic
- Gershon, R. K. 1974. T cell suppression. *Contemp. Top. Immunobiol.* 3:1
- Möller, G., ed. 1977. *Immunology and Differentiation. Immunologic Reviews* #33. Copenhagen: Munksgaard. 145 pp.
- Moretta, L., Ferrarini, M., Mingari, M. C., Moretta, A., Webb, S. R. 1976. Subpopulations of human T cells identified by receptors for immunoglobulins and mitogen responsiveness. *J. Immunol.* 117:2171
- Paul, W. E., Benacerraf, B. 1977. Functional specificity of thymus dependent lymphocytes. *Science* 195:1293
- Shou, L. S., Schwartz, A., Good, R. A. 1976. Suppressor cell activity after concanavalin A treatment of lymphocytes from normal donors. *J. Exp. Med.* 143:1100
- Shreffler, D., David, C. S. 1975. The H-2 major histocompatibility complex in the immune region. *Adv. Immunol.* 20:125
- Stanton, T. H., Boyse, E. A. 1977. A new serologically defined locus, Qa1, in the TL-A region of the mouse. *Immunogenetics* 3:525
- Strelkauskas, A. J., Schauf, V., Wilson, B. S., Chess, L., Schlossman, S. F. 1978. Isolation and characterization of naturally occurring subclasses of human peripheral blood T cells with regulatory functions. *J. Immunol.* 120:1278

PROSTATITIS

◆7323

Edwin M. Meares, Jr., M.D.

Department of Urology, Tufts University School of Medicine,
Boston, Massachusetts 02111

Prostatitis is an entity that is common but poorly understood by most physicians, including urologists. Most of the confusion regarding the etiology and clinical significance of prostatitis relates to the nonstandard methods for diagnosis that have caused clinicians to "lump together" as "prostatitis" diseases of variable type and sequelae. Although important questions remain, recent investigations have significantly enhanced our understanding of certain aspects of prostatitis. Proper therapy varies widely according to the underlying cause of the prostatitis; therefore, a careful search for specific etiology warrants special emphasis.

TYPES OF PROSTATITIS

Nonbacterial prostatitis, a disease of uncertain etiology, is the most common form of prostatitis. A less prevalent but more important type is bacterial prostatitis (acute or chronic), caused by specific gram-negative or gram-positive bacterial pathogens. Special forms of bacterial prostatitis include infections caused by the gonococcus or by the tubercle bacillus. Rarer forms include prostatic infections due to various fungi, nonspecific granulomatous prostatitis, and "allergic" granulomatous prostatitis. Prostatic infections due to other infectious agents, such as viruses, *Trichomonas vaginalis*, T-mycoplasma (*Ureaplasma urealyticum*), or *Chlamydia trachomatis*, have been suspected but remain unproved. This review deals with the two most common varieties of prostatitis: bacterial and nonbacterial prostatitis.

METHODS OF DIAGNOSIS

Although histologic examination of prostatic tissue is generally required to confirm the diagnosis of the unusual forms of prostatitis, such as nonspecific

granulomatous prostatitis, prostatic biopsy is not specifically diagnostic of chronic bacterial prostatitis, since conditions other than bacterial infection produce a similar histologic appearance of tissue reaction to injury. The findings of x ray, cystoscopy, or even rectal examination of the prostate cannot differentiate the most common types of chronic prostatitis. Likewise, since both chronic bacterial and nonbacterial prostatitis are usually associated with the findings of greater than 10 white blood cells per high power field and numerous lipid-laden macrophages (oval fat bodies) in the prostatic expressate, microscopic examination of the expressed prostatic secretions is not specifically diagnostic (1).

Recent investigations have shown that prostatic infections can lead to significant secretory dysfunction of the gland and alterations in the composition of prostatic fluid. For example, Anderson & Fair (2) studied a group of patients with chronic bacterial prostatitis and found a marked decrease in mean levels of specific gravity and zinc, spermine, and cholesterol concentrations in prostatic fluid. They also found a marked increase in mean pH as compared with groups of normal men, men with benign prostatic hyperplasia, and men with prostatic adenocarcinoma. While knowledge of various aspects of prostatic secretory functions is becoming increasingly important in the practice of urology and andrology, such knowledge offers the clinicians no diagnostic specificity regarding prostatitis.

The preferred method for diagnosis of prostatitis is identification of the specific infectious pathogen by means of microbial culture. Because of possible contamination by urethral organisms of nonprostatic origin, isolated culture of the prostatic expressate can be highly misleading in diagnosis. The most accurate confirmation of bacterial prostatitis is the performance of essentially simultaneous quantitative bacteriologic cultures of the urethra, bladder urine, and expressed prostatic secretions (3).

DIAGNOSTIC BACTERIOLOGIC LOCALIZATION TECHNIQUES

Specimen Collection

The voided urine and expressed prostatic secretions are partitioned into segments: the first 10 ml voided [voided bladder 1(VB1)]; the midstream aliquot (VB2); the prostatic secretions expressed by prostatic massage [expressed prostatic secretions (EPS)]; and the first 10 ml voided immediately after prostatic massage (VB3).

The patient should have a full bladder and a genuine desire to void. Skin preparation is generally unnecessary for the circumcised male. The uncircumcised male retracts his foreskin and maintains full retraction throughout collection of all specimens. The glans is cleansed with an antiseptic, such

as povidone-iodine (Betadine®), washed with a sterile wet sponge, and dried. The first 10 ml of voided urine are collected as the patient urinates directly into a sterile tube; after about 200 ml have been emptied from the bladder, the midstream aliquot (VB2) is collected. The patient stops voiding and bends forward. As the physician systematically massages the prostate, he collects drops of prostatic fluid (EPS) directly into a wide-mouthed container. At the end of the massage, the physician exerts gentle stripping pressure on the patient's bulbar urethra to expel into the container any fluid that has pooled in the urethra during the massage. The patient immediately voids again and the first 10 ml of urine voided (VB3) are collected in a fashion similar to that used for the VB1. Even if insufficient EPS has been obtained for culture purposes, a cloudy VB3, when compared to the VB1, usually means that it contains prostatic fluid.

Quantitative Culture Techniques

All specimens are refrigerated immediately until culture techniques are performed. One-tenth milliliter of each specimen is surface streaked onto both blood agar and a differential medium, such as MacConkey or eosin-methylene-blue (EMB) agar. After incubation for 24–48 hours, bacterial colonies are counted and multiplied by 10 to give the quantitative count of bacterial colonies per milliliter on each plate. Standard bacteriologic methods of organism identification are then used.

Interpretation of Cultures

When the bladder urine (VB2) is sterile or nearly so, pathogenic bacteria can usually be localized to the urethra or prostate by a comparison of the bacterial counts of the urethral and prostatic specimens. In urethral infection, the VB1 count significantly exceeds (by at least one logarithm) the counts of the EPS and VB3 cultures. In bacterial prostatitis, the counts of the EPS and VB3 cultures should significantly exceed (by at least one logarithm) the count of the VB1 culture.

When the bladder culture (VB2) reveals significant bacteriuria, the site of infection cannot be accurately localized with these methods. Instead, therapy is instituted with a drug such as nitrofurantoin, penicillin G, or ampicillin, which will usually sterilize the urine but not the prostatic secretions. The segmented cultures are then done 3–4 days later. Despite bactericidal levels of antimicrobial drug in the urine, the prostatic pathogen will still grow on the surface of the agar plates of the EPS and VB3 specimens in cases of prostatitis. Indeed, the most diagnostic pattern of bacterial prostatitis is sterile VB1 and VB2 specimens in the presence of positive EPS and VB3 specimens. Clinical examples of these localization cultures are shown in Tables 1 and 2.

Table 1 Duplicate set of localization cultures showing reproducibility of diagnostic pattern of chronic prostatitis in eight patients. In each instance the bacterial colony counts of the prostatic cultures (EPS and VB3) significantly exceed those of the urethral (VB1) and midstream (VB2) cultures.

Patient	Antibiotic	Colonies per ml				Organism
		VB1	VB2	EPS	VB3	
1	Yes	0	0	1,000	0	<i>Enterococcus</i>
1	Yes	20	0	4,000	20	<i>Enterococcus</i>
2	Yes	0	0	7,000	200	<i>P. mirabilis</i>
2	Yes	0	0	10,000	600	<i>P. mirabilis</i>
3	No	1,000	300	100,000	10,000	<i>E. coli</i>
3	Yes	100	0	10,000	1,000	<i>E. coli</i>
4	Yes	0	0	7,000	10	<i>E. coli</i>
4	No	0	0	10,000	600	<i>E. coli</i>
5	Yes	250	20	5,000	400	<i>Klebsiella</i>
5	Yes	20	0	10,000	1,300	<i>Klebsiella</i>
6	No	60	0	1,000	20	<i>E. coli</i>
6	No	640	40	100,000	200	<i>E. coli</i>
7	Yes	50	10	10,000	1,500	<i>E. coli</i>
7	Yes	0	0	5,000	100	<i>E. coli</i>
8	Yes	0	0	600	200	<i>E. coli</i>
		0	0	500	30	<i>P. mirabilis</i>
8	Yes	0	0	400	100	<i>E. coli</i>
		0	0	600	50	<i>P. mirabilis</i>

Table 2 Localization cultures in eight patients with nonbacterial prostatitis showing normal urethral flora in the urethral cultures (VB1) and no evidence of a prostatic pathogen in the prostatic (EPS and VB2) cultures.

Patient	Antibiotic	Colonies per ml				Organism
		VB1	VB2	EPS	VB3	
1	No	3,000	90	1,200	600	Diphtheroids
		500	10	50	20	Micrococcus
2	No	20	0	10	20	Gamma streptococcus
3	No	4,000	280	1,000	950	Diphtheroids
4	No	280	0	110	60	<i>Staphylococcus epidermidis</i>
5	No	1,500	100	800	500	Alpha streptococcus
6	No	10,000	630	1,500	8,000	Gamma streptococcus
		500	20	150	100	<i>S. epidermidis</i>
7	No	1,500	20	500	30	<i>S. epidermidis</i>
		1,000	0	30	100	Alpha streptococcus
8	No	250	30	0	40	Diphtheroids

BACTERIAL PROSTATITIS

The bacterial pathogens responsible for prostatitis are similar in type and incidence to those that cause urinary tract infection: strains of *Escherichia coli* clearly predominate, although infections due to species of *Klebsiella*, *Proteus*, *Enterobacter*, *Pseudomonas*, *Serratia*, and other gram-negative organisms are sometimes seen. Except for *Enterococcus*, prostatitis due to gram-positive bacteria is less frequent and significant than that caused by gram-negative bacteria (1, 3, 4-7).

ACUTE BACTERIAL PROSTATITIS

Acute bacterial prostatitis is characterized by the sudden onset of chills and fever, perineal and low back pain, urinary urgency and frequency, nocturia, dysuria, generalized malaise and prostration, with or without myalgia and arthralgia, and varying degrees of symptoms of bladder outlet obstruction. Rectal examination usually discloses an exquisitely tender, swollen prostate that is partly or totally quite firm and warm to touch.

Because of the characteristic clinical presentation of the disease, the physician should be highly suspicious of the diagnosis from the history and physical findings. Although prostatic massage will produce grossly purulent prostatic fluid that will grow large numbers of the pathogen on cultures, massage should be avoided because of patient discomfort and the risk of bacteremia. Since acute cystitis usually soon accompanies acute prostatitis, the causative agent can generally be identified by culture of the voided urine.

Hospitalization may be required. General supportive measures, such as adequate hydration, analgesics, bed rest, antipyretics, and stool softeners are indicated. If significant urinary retention requires bladder drainage, placement of a punch suprapubic tube under local anesthesia is preferred to use of an indwelling urethral catheter.

Although the diffuse, intense inflammation of acute prostatitis apparently allows many drugs to diffuse readily into prostatic fluid that normally are excluded, we prefer to initiate antibacterial therapy with trimethoprim-sulfamethoxazole (TMP-SMX), two regular tablets (80 mg TMP and 400 mg SMX) or one double-strength tablet (160 mg TMP and 800 mg SMX) twice daily until the results of the culture and sensitivity testing are known. If the pathogen is sensitive, and if the clinical response is favorable, we continue therapy with this dosage for at least 30 days in an effort to prevent chronic prostatitis from evolving.

As an alternative, initial therapy with gentamicin or tobramycin, 3-5 mg/kg/day divided into three intramuscular doses, plus ampicillin, 2 g given intravenously every six hours, is recommended until the results of the

culture and sensitivity testing are known. A suitable antimicrobial agent should be substituted after about one week and given in full dosage for at least 30 days.

CHRONIC BACTERIAL PROSTATITIS

Although it varies considerably in clinical manifestation, chronic bacterial prostatitis is one of the most common causes of relapsing urinary tract infection in men (1). History of a prior bout of acute prostatitis is often absent. Some men present only because asymptomatic bacteriuria is incidentally found; however, most complain of varying degrees of irritative voiding symptoms, including dysuria, urinary urgency and frequency, nocturia, and pain perceived in various sites (suprapubic, perineal, low back, scrotal, or even in the inner thighs). Some experience postejaculatory pain and there may be intermittent hematospermia. Chills or fever are unusual, unless an acute exacerbation of a chronic infection occurs. Rectal examination discloses no specific or characteristic finding: a tender or "boggy" prostate is not diagnostic. Likewise, urography and cystoscopy reveal no specifically diagnostic findings.

The hallmark of chronic bacterial prostatitis is relapsing urinary tract infection caused by the same pathogen. The organism persists unaltered in prostatic fluid during therapy with the usual antimicrobial agents, because most of these drugs diffuse poorly from plasma into prostatic fluid. Although the urine may be sterilized and the symptoms controlled during therapy, cessation of therapy will eventually result in reinfection of the urine and recurrence of symptoms. Diagnosis is confirmed by performance of bacteriologic localization cultures, as previously outlined.

Antibacterial Therapy and Drug Diffusion

Clinical experience shows that chronic bacterial prostatitis is seldom cured permanently by means of antibacterial therapy. Experiments in dogs (8-17) demonstrate that, except for trimethoprim, most antibacterial agents that are normally useful against the gram-negative bacteria diffuse poorly into prostatic fluid (Table 3). Although erythromycin, oleandomycin, and clindamycin achieve significant levels in prostatic secretions, they are generally ineffective agents against the gram-negative organisms that cause bacterial prostatitis. Chloramphenicol diffuses at a fraction of its plasma level, which limits its usefulness in chronic prostatitis. Regardless of the concentration of the antibacterial agent achieved in prostatic tissue, clinical experience indicates that cure of chronic bacterial prostatitis correlates best with the level of drug achieved in the prostatic fluid.

Table 3 Diffusion of antimicrobial drugs into dog prostatic fluid

High levels	Medium to low levels	Low to negligible levels	
Trimethoprim	Chloramphenicol	Ampicillin	Oxytetracycline
Clindamycin	Lincomycin	Cephalothin	Penicillin G
Erythromycin		Cephalexin	Polymyxin B
Oleandomycin		Doxycycline	Rifampin
		Kanamycin	Rifamide
		Minocycline	Most sulfonamides
		Nitrofurantoin	Tetracycline

Based upon the current theories regarding the mechanisms that govern nonionic diffusion of drugs across biologic lipid membranes (such as prostatic epithelium) and upon the results of animal research, the observation that trimethoprim-sulfamethoxazole is the most successful curative treatment in cases of chronic bacterial prostatitis is not surprising. In our studies of men with chronic bacterial prostatitis who took two tablets of TMP-SMX twice daily for 12 weeks, we found that about 75% were definitely improved and 32% were permanently cured (18), which was twice the rate of cure achieved by similar therapy given for only two weeks (19). Drach (20) reported a cure rate of 33% among patients with chronic bacterial prostatitis who received two tablets of TMP-SMX twice daily for 28 days. Additional studies are needed to determine the optimal dosage and treatment schedule with this medication.

For those patients who are not cured, control of symptoms and protection against urinary tract infection can generally be achieved by means of suppressive therapy given for an indefinite period. Our preference for such therapy is TMP-SMX, one tablet daily, or nitrofurantoin, 50–100 mg daily or twice daily.

Surgical Treatment

While total prostatectomy should be curative, the sequelae of this operation generally preclude its consideration as a treatment alternative in cases of chronic prostatitis. Transurethral prostatectomy can be curative, but only if all foci of infected tissue are successfully removed. In 1976, Smart, Jenkins & Lloyd (7) reported a 40–50% success rate among their patients with chronic prostatitis who underwent “radical transurethral resection”; unfortunately, proof of bacterial prostatitis was not established in most instances. Other investigators (1, 21) found that only about one third of their patients who underwent transurethral prostatectomy for chronic bacterial prostatitis were cured.

NONBACTERIAL PROSTATITIS

Nonbacterial prostatitis, a disease of uncertain etiology, is the most common type of prostatitis seen today. Most often the clinical symptoms and microscopic appearance of the prostatic expressate are quite similar to those found in the patient with chronic bacterial prostatitis; however, the patient with nonbacterial prostatitis seldom (if ever) has a urinary infection documented by culture and characteristically has no positive localization cultures suggestive of an infectious pathogen.

Since the normal microflora of the male urethra often contains small numbers of gram-positive, nonpathogenic "skin inhabitants," such as micrococci, coagulase-negative staphylococci, streptococci, and diphtheroids, one must be cautious in considering these bacteria pathogens during interpretation of localization cultures (Table 2). Much controversy still exists regarding the role of gram-positive organisms in chronic prostatitis; indeed, some still believe these to be the most common prostatic pathogens (21). Careful longitudinal studies of patients who have only gram-positive bacteria (other than enterococcus) on localization cultures, however, usually show a lack of reproducible patterns proving gram-positive prostatitis and no tendency for these organisms to cause urinary tract infection. Since these findings differ significantly from those observed in patients who have gram-negative prostatitis, we agree with others who are confident that chronic prostatitis due to gram-positive bacteria other than enterococcus is uncommon (1, 3, 4-7).

A search for other types of infectious agents in cases of nonbacterial prostatitis has seemingly excluded such pathogens as fungi, obligate anaerobic bacteria, trichomonads, T-mycoplasma (*Ureaplasma urealyticum*), *Chlamydia trachomatis* (TRIC agent), and various viral agents, including cytomegalovirus, *Herpesvirus hominis*, rubella virus, mumps virus, varicella-zoster virus, adenovirus, myxovirus, or paramyxovirus (4, 5, 7, 22-25).

Since the cause of nonbacterial prostatitis is unknown, definitive therapy is difficult. If the clinician suspects that an infectious agent, such as *U. urealyticum* or *C. trachomatis*, may be responsible for the prostatitis, a clinical trial with minocycline or erythromycin in full dosage is a reasonable initial approach to treatment. Unless objective evidence of favorable response to such therapy evolves, however, further therapy with antimicrobial drugs is generally contraindicated. Because symptoms vary and response to any given therapy is also variable, specific treatment should generally be tailored to the individual case. General supportive therapy, emphasizing patient counseling and reassurance, is strongly advocated. Dietary and sexual restrictions are seldom required. During acute symptomatic episodes, relief can sometimes be achieved by administration of propantheline

bromide 15 mg orally four times daily, or oxyphenbutazone, 100 mg orally three times daily for 7–10 days. Therapy with a short course of oral steroids is used only as a last resort. The liberal use of hot sitz baths is generally beneficial, especially during symptomatic flareups. We do not recommend routine use of “therapeutic prostatic massage.”

Not infrequently the clinician encounters the male patient who complains of symptoms of prostatitis, especially a “painful prostate,” but who has negative cultures and no evidence of an inflammatory reaction in his prostatic expressate. Urodynamic evaluation often reveals varying degrees of bladder-sphincter dyssynergia that is responsive to therapy with one of the alpha blockers, such as phenoxybenzamine. Rather than prostatitis, such patients are now being diagnosed as having prostatodynia (26).

CONCLUDING REMARKS

Since the proper management of prostatitis varies considerably according to its etiology, specific diagnosis warrants special emphasis. Chronic bacterial prostatitis, caused primarily by gram-negative bacteria, is one of the most common causes of relapsing urinary tract infection in men. The condition is difficult to cure because most antimicrobial agents diffuse poorly from plasma into prostatic fluid as long as the lipid membrane barrier of prostatic epithelium is intact. Trimethoprim, currently available as trimethoprim-sulfamethoxazole, can be curative, although the optimal dosage and duration of therapy remain uncertain. Nonbacterial prostatitis is the most common and puzzling form of prostatitis. Although the symptoms and findings of a purulent prostatic expressate by microscopic examination clearly suggest an inflammatory disease, the etiology remains obscure. Since the etiology is unknown, cure is difficult to achieve and therapy is usually directed primarily toward control of symptoms and reassurance of the patient.

Literature Cited

1. Meares, E. M., Jr. 1975. Prostatitis: A review. *Urol. Clin. North Am.* 2:3–27
2. Anderson, R. U., Fair, W. R. 1976. Physical and chemical determinations of prostatic secretion in benign hyperplasia, prostatitis, and adenocarcinoma. *Invest. Urol.* 14:137–40
3. Meares, E. M., Stamey, T. A. 1968. Bacteriologic localization patterns in bacterial prostatitis and urethritis. *Invest. Urol.* 5:492–518
4. Meares, E. M. Jr. 1973. Bacterial prostatitis versus “prostatosis”: A clinical and bacteriological study *J. Am. Med. Assoc.* 224:1372–75
5. Mardh, P.-A., Colleen, S. 1975. Search for uro-genital tract infections in patients with symptoms of prostatitis. *Scand. J. Urol. Nephrol.* 9:8–16
6. Fritjofsson, A., Kihl, B., Danielsson, D. 1974. Chronic prostatovesiculitis. Incidence and significance of bacterial findings. *Scand. J. Urol. Nephrol.* 8: 173–78
7. Smart, C. J., Jenkins, J. D., Lloyd, R. S. 1976. The painful prostate. *Br. J. Urol.* 47:861–69
8. Dunn, B. L., Stamey, T. A. 1967. Antibacterial concentrations in prostatic

- fluid. I. Nitrofurantoin. *J. Urol.* 97: 505-7
9. Winningham, D. G., Nemoy, N. J., Stamey, T. A. 1968. Diffusion of antibiotics from plasma into prostatic fluid. *Nature* 219:139-43
10. Winningham, D. G., Stamey, T. A. 1970. Diffusion of sulfonamides from plasma into prostatic fluid. *J. Urol.* 104:559-63
11. Reeves, D. S., Ghilchick, M. 1970. Secretion of the antibacterial substance trimethoprim in the prostatic fluid of dogs. *Br. J. Urol.* 42:66-72
12. Hessel, J. M., Stamey, T. A. 1971. Passage of tetracyclines across epithelial membranes with special reference to prostatic epithelium. *J. Urol.* 106: 253-56
13. Robb, C. A., Carroll, P. T., Tippet, L. O., Langston, J. B. 1971. The diffusion of selected sulfonamides, trimethoprim, and diaverdine into prostatic fluid of dogs. *Invest. Urol.* 8:679-85
14. Granato, J. J. Jr., Gross, D. M., Stamey, T. A. 1973. Trimethoprim diffusion into prostatic and salivary secretions of the dog. *Invest. Urol.* 11:205-10
15. Stamey, T. A., Bushby, S. R. M., Bragonje, J. 1973. The concentration of trimethoprim in prostatic fluid: Non-ionic diffusion or active transport? *J. Infect. Dis.* 128: Suppl., pp. 686-90
16. Reeves, D. S., Rowe, R. C. G., Snell, M. E., Thomas, A. B. W. 1973. Further studies on the secretion of antibiotics in the prostatic fluid of the dog. In *Urinary Tract Infection*, ed. W. Brumfitt, A. W. Asscher, pp. 197-205. London: Oxford Univ. Press. 258 pp.
17. Fair, W. R. 1974. Diffusion of minocycline into prostatic secretion in dogs. *Urology* 3:339-44
18. Meares, E. M. Jr. 1975. Long-term treatment of chronic bacterial prostatitis with trimethoprim-sulfamethoxazole. *Can. Med. Assoc. J.* 112: Suppl., pp. 22-25
19. Meares, E. M., Jr 1973. Observations on activity of trimethoprim-sulfamethoxazole in the prostate. *J. Infect. Dis.* 128: Suppl., pp. 679-85
20. Drach, G. W. 1974. Trimethoprim/sulfamethoxazole therapy of chronic bacterial prostatitis. *J. Urol.* 111:637-39
21. Drach, G. W., Kohnen, P. W. 1977. Prostatitis. In *Urologic Pathology: The Prostate*, ed. M. Tannenbaum, pp. 157-70. Philadelphia: Lea & Febiger. 419 pp.
22. Colleen, S., Mardh, P.-A. 1975. Effect of metacycline treatment on non-acute prostatitis. *Scand. J. Urol. Nephrol.* 9:198-204
23. Gordon, H. L., Miller, D. H., Rawls, W. E. 1972. Viral studies in patients with non-specific prostatic urethritis. *J. Urol.* 108:299-300
24. Nielson, M. L., Justesen, T. 1974. Studies on the pathology of prostatitis. A search for prostatic infections with obligate anaerobes in patients with chronic prostatitis and chronic urethritis. *Scand. J. Urol. Nephrol.* 8:1-6
25. Nielson, M. L., Vestergaard, B. F. 1973. Virological investigations in chronic prostatitis. *J. Urol.* 109:1023-25
26. Drach, G. W., Fair, W. R., Meares, E. M. Jr., Stamey, T. A. 1978. Classification of benign diseases associated with prostatic pain: Prostatitis or prostatodynia? *J. Urol.* 120:266

REGRESSION OF ATHEROSCLEROSIS

◆7324

*Donald W. Crawford, M.D. and David H.
Blankenhorn, M.D.*

Department of Medicine, Cardiology Division, University of Southern California
School of Medicine, Los Angeles, California 90033

INTRODUCTION

In the past, physicians generally felt that atherosclerosis and its complications were an inevitable part of the aging process. However, an alternative concept discussed here, arrest or reversion of the process, is not really new. According to Wilens (1), both Krylov (1916) and Anitschkow (1928) noted that persistence of experimental atherosclerosis in rabbits was dependent upon continued feeding of a cholesterol-rich diet. In 1924, Aschoff (2) suggested that human atherosclerosis might be a reversible disease, noting that the reduced incidence toward the end of the first World War seemed attributable to reduced dietary fat. Following the second World War, Malmros (3) reviewed the history of calorie and fat intake in Scandinavia. In Finland, Norway, and Sweden a reduction in mortality paralleled a reduction of dietary calories in the form of eggs and butter. Strom & Jensen (4) also reported similar findings for Norway. In Denmark, total fat consumption declined but butter and egg consumption increased, and no fall was noted. The concept that dietary lipid, particularly as it effects blood cholesterol level, is a central factor in the production of atherosclerosis, gained further support from population studies of Keys et al (5, 6) and risk factor analyses in Framingham (7).

Unfortunately, trials of diet for primary or secondary prevention of atherosclerotic events have led to conflicting or equivocal reports. Many factors appear reasonable, but we discuss only two: the multiple events contributing to lesion formation and the inherent limitations of human studies where the usual endpoint is clinical perception of a change in the rate of appearance of complications of atherosclerosis. These studies cannot

distinguish between arrest of the disease, regression of the disease, and progression short of that which causes a new complication.

It appears increasingly likely that a common origin of atherosclerosis is endothelial injury, followed by platelet deposition and subintimal cellular proliferation, in addition to incorporation of lipid into the wall. Interactions between hydrodynamics, platelet deposition and release, hormonal effects, mural hypoxia, and metabolic state are not yet fully understood, but all can be involved in the development of early lesions. This suggests many possible avenues of therapeutic approach in addition to diet, blood lipid, or "risk factor" control. Blockade of a single aspect of the complex process might prove effective in some patients. Approaches of this sort cannot be tested adequately in population studies where death or end organ damage serve as end points.

We review briefly here the existing evidence for regression of atherosclerosis in animal models and in humans, in light of present knowledge of factors responsible for the atherosclerotic lesion. Other recent reviews pertaining to the reversibility of the atherosclerotic lesion have been published by Gresham (8) and Wissler & Vesselinovitch (9, 10).

A REVIEW OF PATHOGENESIS

The classic concept of the progression of the early atherosclerotic lesion is reviewed by Duff & McMillan (11). It is generally accepted that so-called fatty flecks or streaks, which contain foam cells and appear in the intima of the aorta in childhood or adolescence, are the sites of later fully developed atheroma. These areas are flat or only slightly raised, and consist of increased glycosaminoglycan in the intima under the endothelium in which fine droplets of fat are scattered, as well as lipid-laden cells or foam cells. Some lesions may regress, but progression is described as an increase in fibrous proliferation, concentration of foam cells in the deeper regions, and formation of a fibrous cap over the main lipid accumulation. Further development consists of foam cell disintegration, which leads to a lipid mass containing necrotic debris and cholesterol crystals. Coalescence occurs, and calcification, hemorrhage, or ulceration produces a complicated lesion, leading to the end organ damage recognized clinically as atherosclerotic vascular disease.

Subintimal Thickening

In 1951, Wilens (12) reviewed an alternative, perhaps complementary concept—that diffuse subintimal thickening, which occurs early in life, predisposes to the accumulation of lipid and, later, the atherosclerotic plaque. This was based on the finding that early intimal thickening is found in its most advanced state in those areas known to be usual later sites of early

atheroma formation. An increased severity of the process was noted in hypertension. The "preatheroma" was characterized as an area of subendothelial smooth muscle cell proliferation by Haust et al (13) and Geer et al (14). Such areas were not found or were less marked in populations in which atherosclerosis was uncommon.

Intimal Injury

French (15) reviewed the possible importance of intimal injury in atherogenesis. In experimental animals the intima is easily injured by many physical and chemical factors, including radiation and carbon monoxide. When injury occurs, muscle cells penetrate into the subendothelial space and new ground substance, collagen and elastin, forms, probably a repair process. Lipid accumulation does not occur in nonlipemic animals. When hypercholesterolemia is present, however, lipid accumulation occurs selectively at injured sites and leads to an exaggerated tissue response (16). A focal distribution of lesions produced by hypercholesterolemia alone may suggest the occurrence of spontaneous, perhaps hemodynamically generated, intimal injury.

Platelet Aggregation

Downie, Mustard & Rowsell (17) found that pigs develop atherosclerosis with increased dietary fat in the absence of significant hypercholesterolemia. They also noted the focal nature of the disease in patterns suggesting hemodynamic injury, and the tendency for deposition of microthrombi. They suggested that incorporation of platelet elements into the wall led to atherogenesis, and that this would be accelerated in areas of endothelial injury.

Additional support for the role of endothelial cell injury and platelet response was presented recently by Harker et al (18) and by Fuster et al (19). Homocystinemia is known to be associated with an increased incidence of atherosclerosis. In baboons, homocystinemia causes patchy deendothelization and increased platelet consumption followed by the lesions of arteriosclerosis (18). Abnormal platelet consumption and plaque formation were prevented by dipyridamole. Fuster et al found that the platelet defect associated with von Willebrand's disease in pigs prevented atherosclerotic plaque development from a high cholesterol diet, although endothelial damage was present (19).

Endothelial Permeability

Physical injury of the endothelium is not necessarily the only mode of increasing intimal permeability and exposing collagen for platelet interactions (see below). For instance, Robertson & Khairallah (20) noted increased endothelial permeability for low density lipoproteins and other

substances in areas exposed to physiologic concentrations of angiotensin II. The process is further accelerated by an associated contraction and degranulation of circulating platelets after attachment to exposed collagen (21). Whether the lesion progresses or regresses may then be determined by the rate of smooth muscle cell migration and proliferation in the subintimal layers as stimulated by low density lipoprotein (22) or platelet factor (23), in addition to the rates of ingress versus removal of lipid due to the increased endothelial permeability. The production of experimental atherosclerosis by hyperlipidemia without apparent physical injury was tentatively explained in this respect by Ross & Harker (24). Chronic hypercholesterolemia led to endothelial cell desquamation and decreased platelet survival as well as lipid accumulation within the wall.

Lipid and Metabolite Exchange

Geer et al (25) suggested that unesterified cholesterol enters the wall by exchange with plasma lipoprotein and that esterification of free cholesterol within the arterial wall is a detoxification mechanism. With this in mind, the physical role of the avascular subintimal thickening due to myocyte proliferation must be considered. Relatively immense diffusional distances are produced, leading to the very real possibility that small molecules such as oxygen are in short supply in the inner avascular layers of the thickened wall (26, 27). Since ample amounts of oxygen are needed for many cellular processes, and possibly esterification of cholesterol, a vicious cycle so generated may progress to the complex lesions associated with clinical vascular disease. Other factors may be associated with the balance between cholesterol ingress and removal. It was recently proposed that high density lipoprotein (HDL) is associated with clearance of cholesterol from the arterial wall (28). Reduced HDL would then increase risk of vascular disease by limiting cholesterol transport from the wall.

REGRESSION OF EXPERIMENTAL ATHEROSCLEROSIS

Diet

In monkeys, relatively short-term feeding of cholesterol with or without other dietary supplements results in an intimal lesion similar to the human fatty streak (29-32). Some months after discontinuation of the dietary supplements, the lesions appear to regress, i.e. lipid content is reduced as observed by histologic examination and chemical analysis. Regression may not occur for some months after removal from the high cholesterol diet (33) and reendothelization of the intima is apparently necessary to begin repair and to halt myocyte hyperplasia (34).

When atherogenic diets are fed for a longer period of time, more intense cellular proliferation occurs. During the regression phase on a normal diet, intimal lesions are again depleted of lipid (35, 36). Luminal cross-sectional area may increase (36). Of interest is the finding that atherosclerosis occurs in primates after four to six months on a diet similar in fat content to the North American diet (37).

Advanced or complicated atherosclerosis quite similar to symptomatic human disease was produced in swine by Daoud et al (38) and Fritz et al (39) who used a combination of endothelial injury and high cholesterol diet. At the end of a low cholesterol regression phase lasting 14 months, atheromas, thrombosis, and hemorrhage essentially disappeared and calcified areas appeared smaller, as compared to those in animals sacrificed at the termination of the cholesterol regimen (38). These findings were confirmed by chemical analysis; arteries containing post regression lesions did not differ greatly in composition from normal arteries (39). Similar results were obtained in rhesus monkeys by Vesselinovitch & Wissler (40).

Combined Therapy

Since hypoxia of a portion of the media is possible after sufficient subintimal proliferation has occurred, Kjeldsen et al (41) fed rabbits excess cholesterol while exposing them to ambient or hyperoxic atmospheres for a period of 10 weeks. Hyperoxic animals appeared to develop less marked atherosclerosis than untreated animals. In an extension of this concept, Vesselinovitch et al (42) reported a standard regression experiment in rabbits in which therapeutic options included low fat diet, hyperoxia, and cholestyramine or estrogen. Most marked regression was seen in the group of animals treated with the combination of low fat diet plus hyperoxia and either drug.

In an attempt to modify other components of atherogenesis, Kwak & Lee (43) utilized two antimetabolites, mercaptopurine and hydroxyurea, and pyridinol carbamate (43). Pyridinol carbamate was thought to act as an endothelial cell relaxant and antiplatelet aggregator. During a low cholesterol regression test in rabbits with atherosclerosis, the antimetabolites and pyridinol carbamate both significantly reduced the surface area of atherosclerotic involvement, as compared to low cholesterol diet alone. Of note in this regard, Kramsch & Chan (44) reported suppression of collagen synthesis in atherosclerotic arteries by ethane hydroxydiphosphate (EHDP).

INDIRECT EVIDENCE FOR REGRESSION IN MAN

Several investigators have used indirect methods or semiquantitative angiography in small therapeutic trials. Öst & Stenson (45) gave nicotinic acid to a small group of patients with leg claudication and used oscillometry and

plethysmography as estimators of changes in blood flow. Approximately one third of the patients showed an improvement in both these tests. The authors felt that development of spontaneous collateral could be excluded in three cases, since serial femoral angiograms appeared to show regression of atherosclerosis.

Zelis et al (46) also used plethysmography in patients with claudication and hyperlipoproteinemia. Diet advice and clofibrate were used as therapy and six patients with Type III hyperlipoproteinemia showed marked improvement after three to six months. It was believed that collateral development did not account for the improvement. Two untreated patients showed no change in peak reactive hyperemic blood flow. Redisch et al (47) treated eight patients with pyridinol carbamate and 15 with placebo. All patients had claudication. Seven patients receiving the medication showed "no progression," as examined by venous occlusion plethysmography and nail bed capillary microscopy and confirmed by observation of serial angiograms. All patients on placebo showed progression.

Another more recent report pertaining to hyperlipoproteinemic patients with claudication was published by O'Connor et al (48). Eight patients were given a placebo and six received dietary supervision. A significant deterioration in peak reactive blood flow was seen in the placebo group as compared to the diet group over a 12-month period.

ANGIOGRAPHIC TECHNIQUES TO STUDY REGRESSION

Regression trials place demands on angiography more complex than and quite different from recognition of vascular obstruction or evaluation to plan for its repair. In order to quantitate atherosclerosis in vessels examined at autopsy, pathologists began sometime ago to solve the problems inherent in this process. These problems occur because the disorder can lead to both dilation and constriction of local segments in the same vessel. In addition, a wide variety of lesion types can occur, thus relative degrees of severity must be assigned to each. This is an entirely feasible procedure, but certain conceptual problems need resolution before it can be done. For example, which is more significant, a hemorrhagic or an ulcerated lesion?

In the evaluation of lesions at autopsy, semiquantitative methods, such as planimetry of lesion area are clearly more precise than visual estimates (49). This is also found in evaluating postmortem angiograms. An additional difficulty in angiographic planimetry arises because only two edges of a vessel are usually observed; variations in density of contrast not on an edge are seldom perceived and never visually quantitated. The findings of Robbins et al (50, 51) may be explained partially on this basis. They found

that quantitative angiographic evaluation made with a caliper improved on visual estimates but still underestimated severity of disease as compared to vessel lumen casts.

The clinical angiographer who would measure lesion change must be aware of other conceptual and practical problems not faced by the pathologist. One of these is vasospasm. Recent attention has been given to the difficulty in differential recognition of transient spasm, misinterpreted as atherosclerotic obstruction (52). Experimental pharmacologic maneuvers may eventually resolve this quandary in part (52). Utilization of the properties of circular symmetry as measured by quantitative angiographic densitometry may also help to resolve the issue (53). In addition to vasospasm, thrombi present a problem because they cannot always be distinguished from plaques by visual inspection of films. Dissolution of a thrombus would mimic rapid plaque resolution. It is possible, however, that advances in the statistical evaluation of fine angiographic lesion characteristics may assist in resolution of this problem, since some progress has been made in the angiographic recognition of lesion types (54).

In order to develop angiographic procedures suitable for regression studies, we began a quantitative angiographic/pathologic study of human arterial disease in 1971 (55). Human cadavers were injected at arterial pressure with a radiopaque silicone rubber bolus, which produces films indistinguishable from those of clinical studies. After catalysis of the silicone rubber, arterial specimens were available for comparison with the three-dimensional information of the cast and with the angiogram. Lesions were assessed with digital densitometric image dissection of the angiograms coupled with automated computer edge-finding and edge roughness evaluation. Pathologic specimens were ordered by a panel with respect to type and extent of disease and used iteratively to "teach" the computer to recognize small degrees of atherosclerosis. The method has proven statistically accurate in evaluation of nonobstructive disease within the range of our cadaver experience. Because of the difficulty in quantitative visual or computerized angiographic evaluation of small, rapidly moving arteries such as the coronaries, we chose to do this study in the femoral artery.

Evaluation of edge contours plays a central role in the computer procedure we developed and is the mainstay of evaluation by human film readers. This cannot reflect in toto the amount or degree of atherosclerosis present even in biplane angiograms. It does not draw an adequate sample of the process because atherosclerosis is not randomly distributed, as noted by Flaherty et al (56) and shown in our studies of high resolution arterial casts (57). To develop a method capable of lesion evaluation in any position on the vessel we studied the profile of density contours within the boundaries of the angiographic shadow (53). Quantitative densitometric information

from an angiogram can yield chord length distributions of angiographic vascular cross sections to an accuracy of about 300 microns. In the presence of disease visible at necropsy, arteries deviated from the regular smooth contour typical of normal arteries. This method is now being computerized and should prove useful not only in more exact evaluation of the extent of disease and its sequential change, but also for most practical purposes the exclusion of vasospasm.

ANGIOGRAPHIC EVIDENCE FOR HUMAN REGRESSION

The literature now reports several cases of arterial obstruction resolution revealed by angiography. For example, Basta et al (58) described a 49-year-old woman with angina, hypertension, hyperlipidemia, and renal artery stenosis as perceived by arteriogram. Hyperlipidemia and hypertension were treated. Four years later, hypertension was absent and renal arteriograms showed a marked reduction in renal artery stenosis.

Starzl et al (59) reported the case of 12-year-old girl with hyperlipoproteinemia who suffered a massive myocardial infarction. A portacaval shunt was done in an attempt to lower blood cholesterol level. Prior to operation, coronary arteriography showed severe obstructive disease. After operation, her cholesterol level declined markedly, and 16 months later the coronary arteries appeared far more patent. Angina pectoris had disappeared.

Case reports, however, have limited applicability to the general population subject to atherosclerotic death. Small pilot regression studies utilizing angiography with various degrees of quantitation have also been published. Knight et al (60) reported 120 patients who had undergone partial ileal bypass in an effort to control hyperlipidemia and atherosclerosis. Angiographic evidence of progression of atherosclerotic disease in these patients is usual, but they believed that the rate of progression was reduced. They reported apparent regression in three patients. However, in a secondary trial using clofibrate in patients, many of whom have received coronary bypass operations, Cohn et al (61) reported no significant influence of therapy. The study of Öst, who employed nicotinic acid, was more encouraging.

Blankenhorn (62) reviewed the current experience with sequential femoral arteriograms in the evaluation of atherosclerosis. Femoral angiograms provide better resolution than is obtained in coronary cinearteriography, but it is still apparent that the vast majority of papers to date report either no change or progression of disease in most patients, with regression in a few. If experimental data pertaining to mobility of cholesterol pools are valid, perhaps this is not surprising. Most reported studies describe late

obstructive disease—the only type of atherosclerosis that commonly calls for angiography and that least likely to change favorably in a relatively short time.

Recently, however, it has been possible to obtain quantitative data from a younger group of patients. Barndt et al (63) demonstrated that nonobstructive femoral atherosclerosis can be quantitated in early adulthood in patients with hyperlipoproteinemia. In quantitative angiographic study of 25 such patients under treatment, regression of disease was present in nine and was significantly correlated with reduction in serum triglyceride, cholesterol, and blood pressure (64).

SUMMARY

Many interacting etiologic factors are necessary for the genesis of the atherosclerotic process. Control of the disease may not depend only upon regulation of blood lipid, but also upon a carefully selected control of endothelial integrity, platelet function, and the hyperplastic response of the arterial wall. Regression of the disease cannot be proven clinically unless carefully controlled sequential radiographic studies are done. Recent radiographic studies suggest that regression of human atherosclerosis can and does occur.

Literature Cited

1. Wilens, S. L. 1947. The resorption of arterial atheromatous deposits in wasting disease. *Am. J. Pathol.* 23:793–804
2. Aschoff, L. 1924. *Lectures in Pathology*. New York: Hoeber. 365 pp.
3. Malmros, H. 1950. The relation of nutrition to health. A statistical study of the effect of the war-time on arteriosclerosis, cardiosclerosis, tuberculosis and diabetes. *Acta Med. Scand. Suppl.* 246:137–53
4. Strom, A., Jensen, R. A. 1951. Mortality from circulatory diseases in Norway 1940–1945. *Lancet* 1:126–29
5. Keys, A., Anderson, J. T., Aresu, M., Björck, G., Brock, J. F., Bronte-Stewart, B., Fidanza, F., Keys, M. H., Malmros, H., Poppi, A., Posteli, T., Swahn, B., Del Vecchio, A. 1956. Physical activity and the diet in populations differing in serum cholesterol. *J. Clin. Invest.* 35:1173–81
6. Keys, A., Kimura, N., Kusukawa, A., Bronte-Stewart, B., Larsen, N., Keys, M. H. 1958. Lessons from serum cholesterol studies in Japan, Hawaii and Los Angeles. *Ann. Intern. Med.* 48: 83–94
7. Kannel, W. B., Dawber, T. R., Friedman, G. D., Glennon, W. E., McNamara, P. M. 1964. Risk factors in coronary heart disease. An evaluation of several serum lipids as predictors of coronary heart disease. The Framingham Study. *Ann. Intern. Med.* 61: 888–99
8. Gresham, G. A. 1976. Is atheroma a reversible lesion? *Atherosclerosis* 23: 379–91
9. Wissler, R. W., Vesselinovitch, D. 1976. Studies of regression of advanced atherosclerosis in experimental animals and man. *Ann. NY Acad. Sci.* 275: 363–78
10. Wissler, R. W., Vesselinovitch, D. 1977. Regression of atherosclerosis in experimental animals and man. *Mod. Concepts Cardiovasc. Dis.* 46:27–32
11. Duff, G. L., McMillan, G. C. 1951. Pathology of atherosclerosis. *Am. J. Med.* 11:92–108
12. Wilens, S. L. 1951. The nature of diffuse intimal thickening of arteries. *Am. J. Pathol.* 27:825–39
13. Haust, M. D., More, R. H., Movat, H. Z. 1960. The role of smooth muscle cells

- in the fibrogenesis of arteriosclerosis. *Am. J. Pathol.* 37:377-89
14. Geer, J. C., McGill, H. C., Strong, J. P. 1961. The fine structure of human atherosclerotic lesions. *Am. J. Pathol.* 38:263-87
 15. French, J. E. 1966. Atherosclerosis in relation to the structure and function of the arterial intima, with special reference to the endothelium. *Int. Rev. Exp. Pathol.* 5:253-353
 16. Still, W. J. S., Dennison, S. M. 1967. Reaction of the arterial intima of the rabbit to trauma and hyperlipemia. *Exp. Mol. Pathol.* 6:245-53
 17. Downie, H. G., Mustard, J. F., Rowsell, H. C. 1963. Swine atherosclerosis: the relationship of lipids and blood coagulation to its development. *Ann. NY Acad. Sci.* 104:539-61
 18. Harker, L. A., Ross, R., Slichter, S. J., Scott, C. R. 1976. Homocystine-induced arteriosclerosis. The role of endothelial cell injury and platelet response in its genesis. *J. Clin. Invest.* 58:731-41
 19. Fuster, V., Bowie, E. J. W., Lewis, J. C., Fass, D. N., Owen, C. A. Jr., Brown, A. L. 1978. Resistance to arteriosclerosis in pigs with von Willebrand's Disease. Spontaneous and high cholesterol diet-induced arteriosclerosis. *J. Clin. Invest.* 61:722-30
 20. Robertson, A. L., Khairallah, P. A. 1972. Effects of Angiotensin II and some analogues on vascular permeability in the rabbit. *Circ. Res.* 31:923-31
 21. Robertson, A. L. Jr., Khairallah, P. A. 1973. Arterial endothelial permeability and vascular disease. The "trap door" effect. *Exp. Mol. Pathol.* 18:241-60
 22. Ross, R., Glomset, J. A. 1973. Atherosclerosis and the arterial smooth muscle cell. Proliferation of smooth muscle is a key event in the genesis of the lesions of atherosclerosis. *Science* 180:1332-39
 23. Ross, R., Glomset, J., Kariya, B., Harker, L. 1974. A platelet-dependent serum factor that stimulates the proliferation of arterial smooth muscle cells in vitro. *Proc. Natl. Acad. Sci.* 71:1207-10
 24. Ross, R., Harker, L. 1976. Hyperlipidemia and atherosclerosis. Chronic hyperlipidemia initiates and maintains lesions by endothelial cell desquamation and lipid accumulation. *Science* 193:1094-1100
 25. Geer, J. C., Sharma, H. M., Pangamala, R. V., Cornwell, D. G. 1977. Progression and regression of experimental atherosclerotic lesions. *Adv. Exp. Med. Biol.* 82:434-39
 26. Zemlenyi, T. 1974. Vascular enzymes and the relevance of their study to problems of atherogenesis. *Med. Clin. North Am.* 58:293-321
 27. Back, L. H. 1976. Analysis of oxygen transport in the avascular region of arteries. *Math. Biosci.* 31:285-306
 28. Miller, G. J., Miller, N. E. 1975. Plasma-high density-lipoprotein concentration and development of ischaemic heart-disease. *Lancet* 1:16-19
 29. Maruffo, C. A., Portman, O. W. 1968. Nutritional control of coronary artery atherosclerosis in the squirrel monkey. *J. Atheroscler. Res.* 8:237-47
 30. Tucker, C. F., Catsulis, C., Strong, J. P., Eggen, D. A. 1971. Regression of early cholesterol-induced aortic lesions in Rhesus monkeys. *Am. J. Pathol.* 65:493-514
 31. Eggen, D. A., Strong, J. P., Newman, W. P. III, Catsulis, C., Malcom, G. T., Kokatnur, M. G. 1974. Regression of diet-induced fatty streaks in Rhesus monkeys. *Lab. Invest.* 31:294-301
 32. Strong, J. P., Stary, H. C., Eggen, D. A. 1977. Evolution and regression of aortic fatty streaks in Rhesus monkeys. *Adv. Exp. Med. Biol.* 82:603-13
 33. Kokatnur, M. G., Malcom, G. T., Eggen, D. A., Strong, J. P. 1975. Depletion of aortic free and ester cholesterol by dietary means in Rhesus monkeys with fatty streaks. *Atherosclerosis* 21:195-203
 34. Weber, G., Fabbrini, P., Capaccioli, E., Resi, L. 1975. Repair of early cholesterol-induced aortic lesions in rabbits after withdrawal from short-term atherogenic diet. Scanning electron-microscopical (SEM) and transmission electron-microscopical (TEM) observations. *Atherosclerosis* 22:565-72
 35. Armstrong, M. L., Megan, M. B. 1972. Lipid depletion in atheromatous coronary arteries in Rhesus monkeys after regression diets. *Circ. Res.* 30:675-80
 36. Armstrong, M. L., Warner, E. D., Connor, W. E. 1970. Regression of coronary atheromatosis in Rhesus monkeys. *Circ. Res.* 27:59-67
 37. Malinow, M. R., McLaughlin, P., Papworth, L., Naito, H. K., Lewis, L., McNulty, W. P. 1976. A model for therapeutic interventions on established coronary atherosclerosis in a nonhuman primate. *Adv. Exp. Med. Biol.* 67:3-31
 38. Daoud, A. S., Jarmolych, J., Augustyn, J. M., Fritz, K. E., Singh, J. K., Lee, K. T. 1976. Regression of advanced

- atherosclerosis in swine. *Arch. Pathol. Lab. Med.* 100:372-79
39. Fritz, K. E., Augustyn, J. M., Jar-molych, J., Daoud, A. S., Lee, K. T. 1976. Regression of advanced atherosclerosis in swine. Chemical studies. *Arch. Pathol. Lab. Med.* 100:380-85
 40. Vesselinovitch, D., Wissler, R. W. 1978. In *State of Prevention and Therapy in Human Arteriosclerosis and in Animal Models*, ed. W. H. Hauss, R. W. Wissler, R. Lehmann, pp. 127-34. Leverkusen: Westdeutscher
 41. Kjeldsen, K., Astrup, P., Wanstrup, J. 1969. Reversal of rabbit atheromatosis by hyperoxia. *J. Atheroscler. Res.* 10:173-78
 42. Vesselinovitch, D., Wissler, R. W., Fisher-Dzoga, K., Hughes, R., Dubien L. 1974. Regression of atherosclerosis in rabbits. Part I. Treatment with low-fat diet, hyperoxia and hypolipidemic agents. *Atherosclerosis* 19:259-75
 43. Kwak, Y. S., Lee, K. T. 1976. Drug-induced regression of experimental atherosclerosis in rabbits. *Fed. Proc.* 35:598 (Abstr.)
 44. Kramsch, D. M., Chan, C. T. 1976. Increased in vivo synthesis of elastin and collagen in atherosclerotic arteries and its suppression by drugs. *Fed. Proc.* 35:598 (Abstr.)
 45. Öst, C. R., Stenson, S. 1967. Regression of peripheral atherosclerosis during therapy with high doses of nicotinic acid. *Scand. J. Clin. Lab. Invest. Suppl.* 99:241-45
 46. Zelis, R., Mason, D. T., Braunwald, E., Levy, R. I. 1970. Effects of hyperlipoproteinemias and their treatment on the peripheral circulation. *J. Clin. Invest.* 49:1007-15
 47. Redisch, W., Clauss, R. H., Rouen, L. R., Katz, M. C., Terry, E. N. 1975. Progression of occlusive atherosclerosis. Long-term administration of pyridinol carbamate. *Arch. Surg.* 110:258-61
 48. O'Connor, J., Ballantyne, D., Pollock, J. G., Cairney, B., Lawrie, T. D. V. 1977. Limb blood flow in treated hyperlipoproteinaemic patients with peripheral vascular disease. *Atherosclerosis* 27:325-31
 49. Cranston, W. I., Mitchell, J. R. A., Russell, R. W. R., Schwartz, C. J. 1964. The assessment of aortic disease. *J. Atheroscler. Res.* 4:29-39
 50. Robbins, S. L., Rodriguez, F. L., Wragg, A. L., Fish, S. J. 1966. Problems in the quantitation of coronary arteriosclerosis. *Am. J. Cardiol.* 18:153-59
 51. McGill, H. C. Jr., Brown, B. W., Gore, I., McMillan, G. C., Pollak, O. J., Robbins, S., Roberts, J. C. Jr., Wissler, R. W. 1968. Grading stenosis in the right coronary artery. *Circulation* 37:460-68
 52. Helfant, R. H. 1978. Coronary arterial spasm and provocative testing in ischemic heart disease. *Am. J. Cardiol.* 41:787-89
 53. Crawford, D. W., Brooks, S. H., Barndt, R. Jr., Blankenhorn, D. H. 1977. Measurement of atherosclerotic luminal irregularity and obstruction by radiographic densitometry. *Invest. Radiol.* 12:307-13
 54. Brooks, S. H., Crawford, D. W., Selzer, R. H., Blankenhorn, D. H., Barndt, R. Jr. 1978. Discrimination of human arterial pathology by computer processing of angiograms for serial assessment of atherosclerosis change. *Comput. Biomed. Res.* In press
 55. Crawford, D. W., Brooks, S. H., Selzer, R. H., Barndt, R. Jr., Beckenbach, E. S., Blankenhorn, D. H. 1977. Computer densitometry for angiographic assessment of arterial cholesterol content and gross pathology in human atherosclerosis. *J. Lab. Clin. Med.* 89:378-92
 56. Flaherty, J. T., Ferrans, V. J., Pierce, J. E., Carew, T. E., Fry, D. L. 1972. In *Atherosclerosis and Coronary Heart Disease*, ed. W. Likoff, B. L. Segal, W. Insull, Jr., J. H. Moyer, pp. 40-83. New York: Grune & Stratton
 57. Crawford, D. W., Barndt, R. Jr., Back, L. H. 1976. Surface characteristics of normal and atherosclerotic human arteries including observations suggesting interaction between flow and intimal morphology. *Lab. Invest.* 34:463-70
 58. Basta, L. L., Williams, C., Kioschos, J. M., Spector, A. A. 1976. Regression of atherosclerotic stenosing lesions of the renal arteries and spontaneous cure of systemic hypertension through control of hyperlipidemia. *Am. J. Med.* 61:420-23
 59. Starzl, T. E., Chase, H. P., Putnam, C. W., Nora, J. J. 1974. Follow-up of patient with portacaval shunt for the treatment of hyperlipidaemia. *Lancet* 2:714-15
 60. Knight, L., Scheibel, R., Amplatz, K., Varco, R. L., Buchwald, H. 1972. Radiographic appraisal of the Minnesota partial ileal bypass study. *Surg. Forum* 23:141-42
 61. Cohn K, Sakai, F. J., Langstron, M. F. Jr. 1975. Effect of clofibrate on progression of coronary disease: a prospective

- angiographic study in man. *Am. Heart J.* 89:591-98
62. Blankenhorn, D. H. 1978. Reversibility of latent atherosclerosis. Studies by femoral angiography in humans. *Mod. Concepts Cardiovasc. Dis.* 47:79-84
63. Barndt, R. Jr., Blankenhorn, D. H., Crawford, D. W. 1974. Prevalence of asymptomatic femoral artery atherosclerosis in hyperlipoproteinemic patients. *Atherosclerosis* 20:253-62
64. Barndt, R. Jr., Blankenhorn, D. H., Crawford, D. W., Brooks, S. H. 1977. Regression and progression of early femoral atherosclerosis in treated hyperlipoproteinemic patients. *Ann. Intern. Med.* 86:139-46

MANAGEMENT OF INSOMNIA

♦7325

*Constantin R. Soldatos, M.D., Anthony Kales, M.D.,
and Joyce D. Kales, M.D.*

Sleep Research and Treatment Center and Department of Psychiatry,
Pennsylvania State University School of Medicine, Hershey, Pennsylvania 17033

INTRODUCTION

The general physician is the most appropriate person to evaluate and treat patients presenting with the complaint of insomnia. However, there has recently been an overemphasis on the need for patients to be evaluated in sleep disorders clinics, particularly in their sleep laboratories. These sleep laboratory evaluations are expensive and time consuming, and often contribute only minimally to the overall management of the insomniac patient. In fact, they may even detract from the comprehensive psychiatric, neurological, pharmacological, and general medical approach that is necessary for the effective management of insomnia. The primary goal of sleep disorders clinics and their staff should be to supplement and enhance the general physician's ability to evaluate and treat patients with sleep disorders, particularly insomnia, the most common sleep disorder (1). This review is designed to strengthen the physician's ability to more rationally and effectively evaluate and treat the condition of insomnia in his office practice. Thus, clinical and laboratory findings are presented, the causative factors of insomnia are defined, assessment guidelines are delineated, and appropriate treatment methods are reviewed.

CLINICAL AND SLEEP LABORATORY FINDINGS

The symptom of insomnia is frequently encountered in medical practice; in a national survey, physicians estimated that nearly one fifth of all their patients present with this complaint (2). The highest percentage (32%) was reported by psychiatrists, which suggests that insomnia is a prominent psychiatric symptom and that psychological factors play a significant role

in its development and persistence. In a survey of representative Los Angeles households, 33% of the respondents had current complaints of difficulty sleeping, and 42% had either current or past complaints of sleep problems (3).

The term insomnia, although literally denoting a complete lack of sleep, is used to indicate a relative inability to sleep that consists of difficulty falling asleep, difficulty staying asleep, early final awakening, or combinations of these complaints. In the Los Angeles survey, when each of the three complaints was considered separately, the prevalence of difficulty staying asleep (23%) far exceeded the prevalence of difficulty falling asleep (14%) or early final wakening (14%). Also, insomnia occurred more often in individuals over 30 years of age as compared to those under 30. When age and sex of the surveyed were considered together, insomnia was found more often than expected in females over 50 years of age (3).

Although difficulty staying asleep is the most common type of insomnia, the single symptom that most frequently marks the onset of insomnia is difficulty falling asleep. In a sample of 211 insomniacs, 32% experienced difficulty falling asleep as their very first symptom, 16% difficulty staying asleep, 5% early final awakening, and the remainder, combinations of these symptoms (4). Difficulty falling asleep also occurred at an earlier age than the other two complaints of insomnia; the mean age of onset was 29 years for difficulty falling asleep, 31 years for difficulty staying asleep, and 37 years for early final awakening.

It is usually the patient with chronic insomnia who comes to the attention of the physician. In these patients, insomnia is quite disabling since prolonged sleeplessness greatly affects their overall well-being. Chronic insomniacs often are refractory to treatment, partially because they are so preoccupied with their sleeplessness that a total evaluation of their condition is severely hampered. They focus on their insomnia and discount or even totally deny the presence of other difficulties, which are usually psychological in nature. The physician, too, is very often almost exclusively concerned with the patient's insomnia, limiting his treatment efforts to prescribing hypnotic medications. In many instances, the use of pharmacologic agents complicates the problem and confuses the clinical picture (5).

Sleep laboratory studies evaluating insomniac patients utilize all-night recordings of EEG, EOG, and EMG (6) as well as several additional physiological measurements. These studies demonstrate a positive correlation between subjective complaints of insomnia and the objective measurements of sleep disturbance in the sleep laboratory (7, 8). Compared to age-matched controls, insomniac subjects show increased objective values for sleep latency, wake time after sleep onset, and total wake time, but the subjective estimates of their sleep difficulty are usually exaggerated. Thus,

the clinical complaint of insomnia is usually valid but often overestimated. In terms of sleep stages, insomniacs have less stage 4 sleep in comparison with controls, whereas their REM sleep is either similar or slightly decreased (9, 10). In addition, insomniac patients show considerable variability from night to night in the values for sleep latency, wake time after sleep onset, stage REM, and stage 4 sleep.

Poor sleepers have higher levels of physiological arousal both prior to and during sleep; heart rate, peripheral vasoconstriction, and rectal temperature are elevated both before and during sleep, and there is an increase in the number of body movements during sleep (9). Another recent finding is a degree of physiological instability in the production of sleep spindles from night to night in insomniacs (11). These findings suggest that the sleep of insomniacs is different qualitatively as well as quantitatively from the sleep of normal subjects.

CAUSATIVE FACTORS

Transient or situational insomnia is a universal but highly individualized phenomenon; almost everyone has experienced insomnia at one time or another (12). Transient sleep difficulties may be caused by stresses created in response to the loss of a loved one, job pressures, concerns regarding examinations, and anticipation or reaction to life changes such as marriage or divorce. Environmental factors such as overcrowding, excessive noise, and poor bedding or housing conditions may also lead to transient sleep difficulty. Insomnia may result from physical restrictions or lack of physical activity. Night-shift work and traveling across time zones (jet lag) are upsetting to internal biological rhythms and often cause transient insomnia. In these latter cases where the sleep cycle has been disrupted, insomnia is associated with daytime sleepiness, psychomotor impairment, anxiety, depression, and various somatic complaints.

Insomnia often accompanies medical disorders that are characterized by pain, physical discomfort, fear, anxiety, or depression (12). Both physical and emotional factors must be considered when insomnia is associated with medical conditions, since there is always an emotional response to a disease process. Medical conditions in which pain and discomfort are primary contributing factors include arthritis, neurogenic pain, various types of headache, and the restless legs syndrome. The pain experienced during the day in these conditions may be greatly magnified when the patients attempt to sleep; there is a decrease in environmental stimuli and consequently more attention is focused internally. In patients with malignancies, in addition to pain being a significant factor, fear and anxiety over the ultimate consequences of their illness are often overwhelming. Patients with angina pecto-

ris and/or cardiac arrhythmias often experience a fear of going to sleep because of a possible attack during the night when they feel the most vulnerable and helpless.

With advancing age, there are distinct changes both in total sleep time and sleep patterns. The elderly not only sleep less and have many nocturnal awakenings, but their sleep is also qualitatively different from the sleep of younger persons (13). The consolidated sleep/wakefulness pattern gradually breaks down with advancing age and a polycyclic pattern begins to reappear. In addition, the aging process itself may contribute to sleep difficulty; medical illnesses with pain and discomfort occur often in the elderly, resulting in frequent sleep disturbance. Also, as the older person faces the realities of declining function and inevitable death, he may become anxious, fearful, and depressed. Fear of death and uncertainty about the "after-death" situation are more prominent during the night, when sleep is perceived as a transient death-like state, and the process of falling asleep involves varying degrees of emotional regression.

It is not uncommon for certain drugs to be the primary cause of insomnia. For example, amphetamines and other stimulants used to suppress appetite, to control narcolepsy, or for other purposes, cause sleeplessness if they are taken too close to bedtime. In addition, steroid preparations frequently result in acute and severe insomnia. Centrally acting adrenergic blockers and various broncho-dilating drugs that are used in the treatment of asthma and that contain ephedrine, aminophylline, and norepinephrine can also result in insomnia. The caffeine contained in coffee and colas may cause sleeplessness; four cups of regular coffee markedly increase sleep latency, whereas four cups of decaffeinated coffee have no effect on sleep latency (14). Finally, an intense degree of insomnia, which we term rebound insomnia, may follow the withdrawal of certain short-acting benzodiazepine hypnotics (triazolam, flunitrazepam, and nitrazepam) (15).

Sleep disturbance is an extremely common complaint of patients with significant psychological conflicts; 70–75% of any given outpatient or inpatient psychiatric population have some type of sleep disturbance (12). The degree of sleep disturbance in schizophrenic patients varies considerably, depending upon whether the process is acute or chronic. Severe sleep disturbances are frequent in acute schizophrenic episodes and often reach the point of total sleeplessness. During the acute manic phase of manic-depressive illness, the total amount of sleep is diminished; insomnia is one aspect of the patient's generalized hyperactivity. Sleep disturbance is one of the most consistent symptoms of depressive illness, with the type of sleep disturbance frequently indicating the type of depression: Endogenous depressives are predominantly troubled by early morning awakenings, whereas reactive depressives usually have more difficulty falling asleep.

In most cases, insomnia is secondary to psychological disturbances. Patients with chronic or severe insomnia usually fall within the diagnostic categories of the neuroses or personality disorders, and have a history of chronic anxiety or depression, or both. About 80% of patients with insomnia have one or more elevated scales on the Minnesota Multiphasic Personality Inventory (MMPI); the most frequently elevated scales are depression, conversion hysteria, psychopathic deviate, and psychasthenia (16). Analysis of the MMPI patterns of insomniac patients strongly suggests that many insomniacs characteristically internalize their emotions rather than express them outwardly.

The internalization of psychological conflicts may lead to chronic emotional arousal, which in turn leads to physiologic arousal and, consequently, insomnia (16). The fact that poor sleepers have higher levels of physiologic arousal during sleep supports this hypothesis. In addition to the underlying psychological disturbances, fear of sleeplessness, which is independent of the primary psychological causes, soon develops in the insomniac patient. Also, a chronic pattern of disturbed sleep eventually conditions the patient to expect insomnia, resulting in a vicious cycle of continued sleep disturbance.

Sleep apnea and myoclonus nocturnus are not significant factors in the development or persistence of insomnia. A prevalence of sleep apnea as high as 5–15% was previously reported for insomniac patients (17). However, in evaluating 150 patients with the primary complaint of insomnia in the sleep laboratory, none exhibited respiratory patterns characteristic of the clinical disorder of sleep apnea (18). Moreover, there was no significant differences between these insomniacs and a group of normal controls similarly evaluated in terms of the number, duration, sleep stage relations, and associated arousal of the apneic episodes during the night. Myoclonus nocturnus was also described as a causative factor in 10–20% of insomnia cases (19). In comparing 106 patients with the primary complaint of insomnia to 78 normal control subjects, the percentage of subjects with nocturnal myoclonic activity in the two groups was very similar: 9% for the insomniacs and 8% for the controls (20). Again, there were no significant differences between the normal and the insomniac groups in the total number of myoclonic events, their mean duration, the number of events preceding EEG arousal, or the total duration of the arousal.

ASSESSMENT

Failure to take an adequate history is one of the most common shortcomings in the management of insomnia (5). A complete history includes a general sleep history, a psychological history, and a drug history. It is

essential to thoroughly describe the sleep problem, including the type(s) of insomnia, the nature of the onset of the problem and its duration (i.e. acute or chronic), the severity and frequency of the sleep difficulty, and the circumstances that precipitate or accentuate the insomnia. The description of the sleep problem should be focused not only on an eight-hour, nightly period, but should assess the sleep-wakefulness patterns for 24 hours.

A detailed history identifying psychological factors and "problems of living" is necessary to properly evaluate the insomniac patient. In this way it is often possible to establish a clear association between the development of certain psychological conflicts and the onset of sleep difficulty. Eliciting this kind of history helps both the patient and the physician to discover any significant psychological conflicts that may underlie the symptom of insomnia. Insomniac patients frequently ignore or are totally unaware of significant marital, sexual, or family problems, which tend to be suppressed and internalized (5).

In addition to obtaining information about drugs that have been prescribed for the patient medically, it is important to learn of any other drugs or agents that the patient has taken. For example, excessive consumption of caffeine can cause an inability to fall asleep, and alcohol can result in difficulty staying asleep. It is common for patients with insomnia to have received inadequate pharmacological treatment, and it is important for the physician to recognize this shortcoming. In the case of antidepressant and antipsychotic medications, the patient is often undertreated; with hypnotic drugs, overtreatment is usually the problem (5).

NONPHARMACOLOGIC TREATMENT

In treating insomnia, the physician must be aware that the disorder is multidimensional: any approach that is directed to only one of the factors involved in insomnia will usually be inadequate and unsuccessful.

It is helpful to increase physical activity and exercise levels during the day, but not close to bedtime since exercise at that time may heighten physiologic arousal (12). Complex mental activity, such as studying, too late in the evening may aggravate an insomniac condition; thus, patients are instructed to avoid mentally stimulating situations and engage instead in relaxing mental activities prior to bedtime. In addition, patients are urged to regulate their daily schedules and to establish a regular bedtime hour, although flexibility of the patient's sleep schedule must be maintained so that he does not become unduly obsessive about the schedule itself. The insomniac should go to bed only when he is sleepy. If he is unable to sleep, he should get up, leave the bedroom and engage in some kind of relaxing activity. In this way, the patient learns to associate the bed with sleep rather

than with obsessive thoughts and concerns. Naps during the day should be totally discouraged. Finally, drinking beverages containing caffeine should also be discouraged, particularly before or close to bedtime (12).

Support, education, and reassurance often help to minimize the patient's fear of sleeplessness and the consequences of insufficient sleep. Patients are often obsessed with the fear that their insomnia will severely affect their physical health and even cause their death. It can be very helpful to explain to the patient how his own anxiety can become part of a vicious cycle that accentuates and maintains insomnia.

Behavioral treatment approaches that minimize rumination and help the patient to focus his attention can be beneficial (21). Focusing one's attention on a particular object, breathing, or counting substitutes attention and patterned thought for the ruminative concerns that otherwise maintain the insomniac's cognitive arousal at high levels. In addition, as in biofeedback and relaxation procedures, the individual's active participation reduces his feelings of passivity and helplessness. However, behavioral therapies for insomnia are usually indicated only for patients who have difficulty falling asleep. In addition, these therapies require a much greater degree of cooperation and compliance by the patient.

Referral for psychiatric treatment can greatly improve the prognosis of the chronic insomniac. Psychodynamic psychotherapy, for example, attempts to resolve the original conflict underlying the insomnia and its development. A chronic insomniac patient may fear going to bed because of suppressed memories of traumatic events and experiences in childhood or adolescence that were associated with sleep or bedtime (e.g. parental drunkenness, domestic violence, or incest).

Sexual problems that may result in repeatedly frustrating sexual relations or total avoidance of sexual relations are common in patients with insomnia (12). In marriages with poor sexual relations the continuous deprivation of affection and closeness can lead to tension and feelings of being neglected, which can in turn result in insomnia. Sexual counseling and conjoint marital therapy are often necessary once a marital or sexual problem has been identified and acknowledged.

In treating the psychological aspects of insomnia, it is generally necessary for the physician to be active and direct in exploring conflict areas rather than utilizing a gradual, uncovering approach. By delineating conflict areas early in treatment, there is an increased likelihood that the patient will become actively involved in therapy. The insomniac's need to defend against a loss of control may result in his attempting to manipulate his sleep medication by varying the drugs, the dosages, or both, contrary to the physician's recommendations (12). The potential for this problem should be explained to the patient in terms of its value as an immediate psychological

insight, and because pharmacological treatment is frequently rendered useless by this improper adjustment of medication.

CHEMOTHERAPY

Over the last several years, the precise and objective measurements of the sleep laboratory have been utilized in the evaluation of the effectiveness of hypnotic drugs, particularly when these medications are administered for more than one to two weeks (22, 23). This long-term evaluation is particularly valuable since over 40% of all prescriptions for sleeping pills are for periods of longer than three months (24). Many hypnotic drugs, including glutethimide (Doriden®), ethchlorvynol (Placidyl®), methaqualone (Parest®), chloral hydrate (Noctec®), pentobarbital (Nembutal®), secobarbital (Seconal®), and flurazepam (Dalmane®) are relatively effective for inducing and maintaining sleep with initial and short-term use (25). However, except for flurazepam there is a considerable loss of effectiveness with intermediate-term use of these drugs (25). Studies have demonstrated that flurazepam continues to be effective with one month of consecutive nightly use, whereas other drugs lose most of their effectiveness within a two-week period. In most patients, treatment should be initiated with a dosage of 15 mg of flurazepam at bedtime. This nightly dosage should be continued for at least one week and usually for two weeks, since this period of time is necessary for the maximal effects of the drug's active metabolites to develop. If after one to two weeks the 15-mg dose does not sufficiently improve sleep, the dosage should be increased to 30 mg.

On the first and second nights following withdrawal of flurazepam, sleep is still significantly improved, whereas the withdrawal of most hypnotic drugs leads to an immediate return of sleep difficulty to pre-drug levels, or in some cases to rebound insomnia. This carry-over effectiveness of flurazepam facilitates both total withdrawal and temporary withdrawal from the drug, such as for the one- or two-day "drug holiday." Prescribing flurazepam for nightly use during the week and allowing for a drug holiday on the weekend, with very little loss of effectiveness on the nondrug nights, permits the patient to develop the attitude that he can sleep without the medication.

As with many prescription drugs, claims that over-the-counter drugs effectively induce and/or maintain sleep have not been verified in the sleep laboratory. In one study, two tablets of Sominex® (methapyrilene HCl, 25 mg, and scopolamine HBr, 25 mg) did not produce any favorable changes in inducing or maintaining sleep (26). Sodium salicylamide, a common ingredient of several over-the-counter preparations in doses of 200–400 mg, lacked hypnotic properties in a higher dose of 650 mg (27). An even higher dose (1350 mg) had some sedative effects, but there was no clear-cut hyp-

notic efficacy. Over-the-counter preparations containing scopolamine not only lack proven effectiveness, but also can be hazardous since even in the recommended dose range they may precipitate acute glaucoma, especially in elderly patients who have a narrow corneal-iris angle. Also, a dose of two to three times the recommended dosage may produce transient disorientation and hallucinations, particularly in emotionally unstable individuals (28).

In treating sleep disturbances that accompany schizophrenia, neuroleptics with sedative effects, such as chlorpromazine (Thorazine®) or haloperidol (Haldol®) are most helpful. Rather than complicating the clinical picture by prescribing a hypnotic agent at bedtime, a larger dose of this neuroleptic drug should be administered so that the sedative side effects will aid in inducing sleep (12).

In treating sleep disturbances associated with retarded endogenous depression, an energizing tricyclic antidepressant such as imipramine (Tofranil®) is used in divided doses during the day together with a benzodiazepine hypnotic such as flurazepam at bedtime. When the therapeutic effects of the tricyclic have been well established, which takes about two to three weeks, flurazepam can be discontinued, since the relief of insomnia usually accompanies the improvement of the underlying depression. In agitated endogenous depression, a tricyclic antidepressant with sedating effects such as amitriptyline (Elavil® or Endep®) may be used for treating both the depressive process as well as the accompanying symptom of insomnia (12).

Attaining a sufficient dosage level is critical to the effectiveness of both antidepressant and antipsychotic medications (5). In the case of the tricyclic antidepressants, many patients receive a total daily dosage of less than 100 mg, a level that is far from adequate for obtaining the basic antidepressant effect. Another problem with the tricyclic antidepressants with sedating side effects such as amitriptyline and doxepin is that patients are prescribed too much of their medication during the day rather than at bedtime. Thus, instead of receiving the benefit of the sedative side effects at bedtime, when they are most needed, the patient must cope with the sedative side effects during the day, when they are very discomforting and may present a hazard in performing critical tasks such as driving an automobile. The sedative side effects of the tricyclic antidepressants are immediate, while the antidepressant effect is more likely to occur after 7–10 days. Confusing the side effects of these drugs with their basic action may account for much of the difficulty in attaining a sufficient total daily dosage, and also in appropriately adjusting the daytime-to-bedtime dosage ratio (5).

In conditions where the insomnia is secondary to neurotic anxiety states, the use of benzodiazepine anxiolytics (e.g. chlordiazepoxide, diazepam) is

indicated both for the daytime anxiety and in an increased bedtime dose for the sleep difficulty. An advantage of this class of drugs, which includes anxiolytics and hypnotics, as opposed to the nonbenzodiazepine hypnotosedatives, is the lack of drug interactions such as a reduction in the effectiveness of coumarin anticoagulants (29). Benzodiazepines are often considered alike in their actions, even though clear-cut differences have been demonstrated in their absorption, distribution, metabolism, and elimination. Such discoveries as the occurrence of rebound insomnia following the withdrawal of short-acting benzodiazepine hypnotic drugs accentuate the need to consider each benzodiazepine compound separately and not simply as part of a class (15).

In all circumstances, hypnotic drug administration is a symptomatic treatment for insomnia and should be used only as an adjunct to the main therapeutic endeavors directed toward the underlying cause of this condition, which is commonly psychological. In medical conditions, symptomatic treatment of the insomnia may be indicated in addition to treatment of the primary disease process. Nevertheless, treatment of the primary medical condition should take precedence over treatment for the insomnia itself in the interest of the patient's benefit and since the insomnia is usually alleviated once the primary condition is effectively treated.

SUMMARY

The general physician is best qualified to evaluate and treat the insomniac patient. In addition to a general medical history and physical examination, the assessment of the insomniac patient should include a complete sleep history, a psychological history, and a drug history. In evaluating and treating insomnia, the physician should keep in mind that the etiology of this disorder is multidimensional. However, the most common factors are psychological in nature.

Support, education, reassurance, and instructions for general measures are an important part of the management of insomnia. Psychodynamic or behavioral psychotherapies may be utilized; conjoint marital therapy and sexual counseling are especially helpful if specific marital or sexual problems exist.

Flurazepam is the only hypnotic that has been demonstrated to be effective beyond one to two weeks of continued nightly administration; it continues to be effective with one month of consecutive nightly use. In all cases, the use of hypnotic medication is only an adjunct to the overall treatment of the insomniac patient.

When insomnia is secondary to depression, a sedative antidepressant should be administered in a single bedtime dose or with the larger dose at

bedtime. The physician should gradually increase the dose to an effective, therapeutic level and not use the sedative effects of the drug in determining the end point in the dosage; the sedative effects of the drug are immediate while the basic antidepressant effect takes one to two weeks to develop.

For the schizophrenic patient with insomnia, a sedative neuroleptic is the drug of choice, whereas for the patient with an anxiety state, a benzodiazepine anxiolytic is indicated. In both cases, a larger dose of the medication is given at bedtime.

In medical conditions, pain, discomfort, fear, anxiety, and depression are frequent components or reactions to the basic condition that often lead to insomnia. In medical disorders complicated with insomnia, treating the primary condition is essential. However, it is emphasized that there are emotional components and reactions to all medical conditions. Thus, the physician should also attempt to assess and alleviate any psychological factors arising from the medical condition as well as providing symptomatic treatment.

Literature Cited

1. Kales, A., Kales, J. D. 1974. Sleep disorders. *N. Engl. J. Med.* 290:487-99
2. Bixler, E. O., Kales, J. D., Scharf, M. B., Kales, A., Leo, L. A. 1976. Incidence of sleep disorders in medical practice: A physicians' survey. *Sleep Res.* 5:160
3. Kales, A., Bixler, E. O., Leo, L. A., Healy, S., Slye, T. 1974. Incidence of insomnia in the Los Angeles metropolitan area. *Sleep Res.* 3:139
4. Bixler, E. O., Kales, J. D., Kales, A., Bianchi, S. B., Fahs, J. J. 1978. Subjective clinical characteristics of insomniac patients. *Sleep Res.* 7. In press
5. Kales, A., Kales, J. D., Bixler, E. O., Martin, E. 1975. Common shortcomings in the evaluation and treatment of insomnia. In *Hypnotics: Methods of Development and Evaluation*, ed. F. Kagan, T. Harwood, K. Rickels, A. Rudzik, H. Sorer, pp. 29-40. New York: Spectrum. 432 pp.
6. Rechtschaffen, A., Kales, A., eds. 1968. *A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects*. Public Health Service Publications No. 204, Washington, DC: GPO. 56 pp.
7. Bixler, E. O., Kales, A., Leo, L. A., Slye, E. C. 1973. A comparison of subjective estimates and objective sleep laboratory findings in insomniac patients. *Sleep Res.* 2:143
8. Carskadon, M. A., Dement, W. C., Mitler, M. M., Guilleminault, C., Zarcone, V. P., Spiegel, R. 1976. Self-reports versus sleep laboratory findings in 122 drug-free subjects with complaints of chronic insomnia. *Am. J. Psychiatry* 133:1382-88
9. Monroe, L. J. 1967. Psychological and physiological differences between good and poor sleepers. *J. Abnorm. Psychol.* 72:255-64
10. Kales, A. 1969. Psychophysiological studies of insomnia. *Ann. Intern. Med.* 71:625-29
11. Soldatos, C. R., Scarone, S., Bixler, E. O., Scharf, M., Kales, A. 1978. Sleep characteristics and phasic physiological events in insomniac patients. *Sleep Res.* 7. In press
12. Kales, A., Kales, J. D., Bixler, E. O. 1974. Insomnia: An approach to management and treatment. *Psychiatr. Ann.* 4:28-44
13. Feinberg, I., Koresko, R., Heller, N. 1967. EEG sleep patterns as a function of normal and pathological aging in man. *J. Psychiatr. Res.* 5:107-44
14. Karacan, I., Thornby, J. I., Anch, A. M., Booth, G. H., Williams, R. L., Salis, P. J. 1976. Dose-related sleep disturbances induced by coffee and caffeine. *Clin. Pharmacol. Therap.* 20:682-89
15. Kales, A., Scharf, M. B. 1978. Rebound insomnia: A new clinical syndrome. *Science* 201:1039-41

16. Kales, A., Caldwell, A. B., Preston, T. A., Healey, S., Kales, J. D. 1976. Personality patterns in insomnia. *Arch. Gen. Psychiatry* 33:1128-34
17. Guilleminault, C., Eldridge, F., Dement, W. C. 1973. Insomnia with sleep apnea: A new syndrome. *Science* 181: 856-58
18. Bixler, E. O., Soldatos, C. R., Scarone, S., Russek, E., Kales, A., Thomson, K. C. 1978. Sleep apneic episodes in insomniac patients and normal subjects. *Sleep Res.* 7. In press
19. Guilleminault, C., Raynal, D., Weitzman, E. C., Dement, W. C. 1975. Sleep-related periodic myoclonus in patients complaining of insomnia. *Trans. Am. Neurol. Assoc.* 100:19-22
20. Bixler, E. O., Soldatos, C. R., Scarone, S., Martin, E. D., Kales, A., Charney, D. S. 1978. Similarities of nocturnal myoclonic activity in insomniac patients and normal subjects. *Sleep Res.* 7. In press
21. Montgomery, I., Perkin, G., Wise, D. 1975. A review of behavioral treatments for insomnia. *J. Behav. Ther. Exp. Psychiatry* 6:93-100
22. Kales, A., Kales, J. D., Bixler, E. O., Scharf, M. B. 1975. Methodology of sleep laboratory drug evaluations: Further considerations. See Ref. 5, pp. 109-26
23. US Food and Drug Administration. 1977. *Guidelines for the Clinical Evaluation of Hypnotic Drugs*. US Department of Health, Education and Welfare (FDA) Publications No. 78-3051, Washington, DC:GPO. 14 pp.
24. Bixler, E. O., Kales, J. D., Kales, A. 1976. Hypnotic drug prescription patterns: Two physicians' surveys. *Sleep Res.* 5:62
25. Kales, A., Bixler, E. O., Kales, J. D., Scharf, M. B. 1977. Comparative effectiveness of nine hypnotic drugs: Sleep laboratory studies. *J. Clin. Pharmacol.* 18:207-13
26. Kales, J. D., Tan, T. L., Swearingen, C., Kales, A. 1971. Are over-the-counter sleep medications effective? All-night EEG studies. *Curr. Ther. Res.* 13:143-151
27. Soldatos, C. R., Kales, A., Bixler, E., Scharf, M. B., Kales, J. D. 1978. Hypnotic effectiveness of sodium salicylamide with short-term use: Sleep laboratory studies. *Pharmacology* 16:193-98
28. Bernstein, S., Lef, R. 1967. Toxic psychosis from sleeping medicine containing scopolamine. *N. Engl. J. Med.* 277: 638-39
29. Greenblatt, D. J., Shader, R. I. 1975. Drug interactions in psychopharmacology. In *Manual of Psychiatric Therapeutics*, ed. R. I. Shader, pp. 269-79. Boston: Little, Brown. 362 pp.

DIAGNOSIS AND MEDICAL MANAGEMENT OF INSULINOMAS

◆7326

Stefan S. Fajans, M.D. and John C. Floyd, Jr., M.D.¹

Division of Endocrinology and Metabolism and the Metabolism Research Unit,
Department of Internal Medicine, University of Michigan, Ann Arbor,
Michigan 48109

INTRODUCTION

The advent of more precise definitions of pancreatic islet physiology and of expanded techniques for its study in man have facilitated a more dependable, simpler, and sometimes earlier diagnosis of "insulinoma." Further, advances have been made in the medical management of insulinoma that benefit patients in whom surgical therapy is not feasible or curative or must be delayed. Therefore, it is appropriate to review current methods of diagnosis and medical management of insulinoma. This discussion concerns itself not only with solitary tumors of the islets of Langerhans, but also with other islet cell disease characterized by inappropriate secretion of insulin and hypoglycemia in the adult. We interpret clinical and laboratory findings and current methods of diagnosis based on knowledge of normal homeostatic control of insulin secretion and aberrations of secretion in pathological states.

Cellular mechanisms by which hyperinsulinism leads to fasting hypoglycemia, an overall classification of spontaneous hypoglycemia, symptoms and signs of hypoglycemia, and the clinical course of patients with hyperinsulinism are not covered since they were recently reviewed (1-3). We briefly discuss hypoglycemia due to nonpancreatic tumors, factitious hypoglycemia, and the insulin autoimmune syndrome, since these entities may pose diagnostic challenges in differential diagnosis of insulinoma.

¹Supported in part by USPHS Grants AM 02244, National Institutes of Arthritis, Metabolism, and Digestive Diseases, NIH; and 5M01-RR-42, General Clinical Research Centers Program of the Division of Research Resources, NIH.

CLASSIFICATION OF PANCREATIC ISLET BETA CELL DISEASE

Among adults with hyperinsulinism, a number of pathologic entities involving islet tissue can be recognized that may be clinically indistinguishable. We have encountered most of these entities among our last 40 patients with organic hyperinsulinism. (a) Single benign adenomas were most common (68%). (b) Scattered microadenomatosis (10%) interspersed with normal islets, without or with one or more macroadenomas is not uncommon in our experience (4), though it is rarely discussed in the literature. The recognition of microadenomatosis is important therapeutically, since removal of only the macroadenomas usually will not cure the patient. In some, a 70% pancreatectomy will achieve a cure, while in others additional medical therapy is essential. (c) Islet cell carcinoma (8%) can be diagnosed only with certainty when regional or distant metastases are present. Since distant metastases, usually in the liver, cannot be removed, medical therapy is essential. (d) In patients with multiple endocrine neoplasia Type I (10%), which usually occurs in families, single adenomas, multiple adenomas, or islet cell carcinoma with regional lymph node metastases were seen. Microadenomatosis has been recognized as well. (e) In the past, nesidioblastosis (diffuse proliferation of the islet cells of the pancreas) was recognized only in infancy. We have seen a 13-year-old girl with a four-year history of symptoms of hypoglycemia and a 47-year-old man with this diagnosis. After 70% pancreatectomy, the child has required medical therapy (diazoxide plus trichlormethiazide), while the adult patient has remained cured at a follow-up of one year. (f) In addition, hyperplasia has recently been reported.

DIAGNOSIS AND DIFFERENTIAL DIAGNOSIS OF INSULINOMA

Laboratory Tests: Essential Diagnostic Tests

The diagnosis of insulinoma or inappropriate hyperinsulinism due to pancreatic islet beta cell disease can be made by determination of fasting plasma levels of glucose and of immunoreactive insulin (IRI) on blood samples obtained after an overnight fast or during prolongation of the overnight fast. Characteristically hypoglycemia occurs on fasting, and the diagnosis can be confirmed by the demonstration of an inappropriately raised plasma level of IRI.

FASTING PLASMA LEVELS OF GLUCOSE In healthy males, plasma glucose levels do not decline below 60 mg/dl during a 24-hour fast. During a

70–72-hour fast, they usually decline further but not below 55 mg/dl (2, 5). Fasting results in greater decreases in blood glucose in healthy premenopausal females than in healthy males (5). Merimee & Tyson (5) reported that in approximately 40% of women, blood glucose levels decreased to less than 45 mg/dl during the 72-hour fast. Of these women, one third had values as low as 30–35 mg/dl, with occasional values between 22 and 30 mg/dl. These authors suggest that after one day of fasting, a feasible lower limit of normal in women would be approximately 35 mg/dl for plasma glucose, and that after 36-hour fasting it is impossible to differentiate physiologic hypoglycemia from that secondary to hyperinsulinism. None of these subjects had symptoms unequivocally due to hypoglycemia. We confirmed that there is a sex difference in plasma glucose levels during 72 hours of fasting (2). However, the lowest level in our females was 36 mg/dl, as compared to approximately 46 mg/dl in another report (6) and to 22 mg/dl in the group of females referred to above (5). In many healthy females in whom plasma glucose falls below 50 mg/dl in the course of a 72-hour fast, the level will rise spontaneously again as the fast proceeds, presumably due to accelerated gluconeogenesis (2).

In most patients with pancreatic islet cell disease, the level of plasma glucose after an overnight fast (10 hours) is usually below 60 mg/dl. However, in approximately one half of our patients with proven islet cell disease, the overnight fasting plasma glucose level was above 60 mg/dl on the first day of the study, and in seven patients it continued so for seven consecutive days. Thus, fasting plasma glucose may fluctuate from normal to subnormal levels from day to day. A convenient way to demonstrate fasting hypoglycemia in such patients is to prolong the overnight fast for four hours or more. This procedure usually will reduce plasma glucose to below 50 mg/dl. When the fast is lengthened to 20 hours, more severe fasting hypoglycemia (plasma glucose below 35–40 mg/dl) will develop in the majority of patients. In men, this is clearly diagnostic of fasting hypoglycemia. According to Merimee & Tyson (5), such levels are nondiagnostic in females. Differentiation between normal and abnormal can be resolved easily by simultaneous measurements of plasma insulin (see below; 2). If after 24 hours of fasting hypoglycemia is not induced, the fast should be continued for up to 70 hours. If symptomatic or biochemical hypoglycemia has not developed, the patient should be exercised vigorously for two hours at that time. In patients with insulinoma and very low grade insulin secretion, or with other types of fasting hypoglycemia, exercise will produce a further fall in blood glucose levels, but in healthy subjects and in patients with functional hypoglycemia it usually raises blood glucose. Exercise at any time during the fast can precipitate or aggravate hypoglycemia in patients with insulinoma.

IMMUNOREACTIVE INSULIN AFTER OVERNIGHT FAST Determination of the concentration of plasma insulin after an overnight fast is essential to establish the presence of hyperinsulinism due to pancreatic islet cell disease. When fasting insulin levels are measured on several days, elevated levels of insulin in peripheral venous plasma usually can be found in at least one specimen. Absolutely elevated levels are not found in all patients with functioning islet cell disease, owing in part to removal by the liver of some of the secreted insulin before it appears in peripheral blood. Thus, although a diagnosis of insulinoma is confirmed by an elevated fasting insulin level, a "normal" level does not rule out this diagnosis. In our laboratory, the normal upper limit of concentration of insulin in peripheral blood of fasting nonobese healthy subjects was $24 \mu\text{U/ml}$ (mean + 2 SD) (1). By this standard, normal levels were found in approximately half of 40 patients on the first day of study, and in six they remained normal through seven consecutive days. More recently, in a larger group of healthy subjects we determined that the mean plus two standard deviations of fasting plasma IRI was $20 \mu\text{U/ml}$. On prolongation of the overnight fast, plasma insulin in males and females declines steadily to reach the range of $4\text{--}7 \mu\text{U/ml}$ from 36 through 70 hours. Thus, as fasting continues, plasma insulin declines with plasma glucose. On the other hand, in the patient with pancreatic islet cell disease, when plasma glucose level falls into the hypoglycemic range on prolongation of the fast, levels of plasma insulin remain constant or rise. This is evidence for the presence of autonomous insulin-producing tissue.

RATIO OF IMMUNOREACTIVE INSULIN TO PLASMA GLUCOSE AFTER OVERNIGHT FAST The relation of plasma insulin to plasma glucose can be expressed as the ratio of the plasma concentration of immunoreactive insulin ($\mu\text{U/ml}$) to the concentration of plasma glucose (G) (mg/dl). In healthy nonobese males, the normal upper limit (mean + 2SD) of IRI/G after an overnight fast was reported by us as 0.30, based on the log of normal distribution of 412 IRI/G values of 114 control subjects (1). These values are based on the particular methods for plasma glucose and plasma insulin used in our laboratory. As healthy subjects undergo prolonged fasting, plasma insulin declines with plasma glucose; the ratio IRI/G remains constant or declines modestly and during the 47–70-hour period becomes similar in both sexes (2).

The IRI/G ratio provides the most reliable data to confirm or establish a diagnosis of insulinoma in patients with mild abnormalities. A plasma insulin level in the "normal" range and an associated fasting plasma glucose level in the hypoglycemic range constitute evidence that the secretion of

insulin is excessive. This is made readily apparent by an elevated IRI/G ratio. In patients with pancreatic islet cell disease in whom fasting plasma glucose (or IRI/G) is within the normal range after a 10-hour fast, prolongation of the fast for four hours or more causes IRI/G to rise into the abnormal range. After 14 hours of fasting, all but two of our 40 patients had abnormal ratios, and after 19 hours all had abnormal ratios. In the course of a 72-hour fast, a slight elevation of the IRI/G ratio at one or two time points should not be interpreted as signifying pancreatic islet cell disease. In patients with such disease the elevations in the ratio are continuous and significant (1, 2, 7).

Because of hyperinsulinemia due to insulin resistance, basal IRI/G in normal obese subjects may be elevated as compared to ratios of healthy nonobese subjects and therefore may simulate the ratios of patients with insulinoma. Since, in such patients, fasting plasma glucose is normal and does not fall below 55 mg/dl on prolonged fasting, there is no problem in differential diagnosis. The pattern of change of the IRI/G ratio at progressive time intervals of fasting provides the best distinction between patients with pancreatic islet cell disease and patients, nonobese or obese, without autonomous hyperinsulinism (1, 2, 7). The numerical value of the IRI/G ratio depends upon the techniques employed for measuring plasma glucose and insulin, but the change in the ratio with time is independent of these variables (7).

Turner et al (8) suggested that the "amended" IRI/G ratio [plasma IRI (μ U/ml) $\times$ 100 divided by plasma glucose (mg/dl) - 30] would better discriminate abnormal ratios. The rationale for subtracting 30 mg/dl from the plasma glucose level is that in healthy subjects at a plasma glucose level of 30 mg/dl, secretion of insulin would be reduced sufficiently so that plasma insulin in peripheral blood would fall in the range of 0-1 μ U/ml in healthy subjects. We found a range of "amended" IRI/G ratios in our healthy control subjects (1) similar to that found by Turner et al in theirs (8). Again, the change in the ratio during more prolonged fasting (ratio rises) gives a better diagnostic confirmation of the presence of inappropriate islet cell secretion (2) than does observation after overnight fasts.

The patient in whom pancreatic islet cell disease is strongly suspected should be studied as an inpatient. On the other hand, if suspicion is not strong, the patient can be studied as an outpatient. We instruct the patient to have his last meal at 5-6 PM the night before and to come to the laboratory in the fasting state at 8 AM the following morning. Hourly plasma samples for determination of glucose and insulin are obtained in the morning and throughout the afternoon unless clear-cut symptoms of hypoglycemia become apparent and the test is terminated. If fasting is prolonged

to 18–22 hours and IRI/G remains in the normal range, the diagnosis of pancreatic islet cell disease with hyperinsulinism is ruled out for practical purposes. In some patients after an overnight fast, pathologically elevated levels of immunoreactive insulin are not associated with a hypoglycemic level of plasma glucose. This may be due to a reduction in the number and/or affinity of insulin receptors observed in patients with insulinoma and may constitute a protective mechanism against excessive hypoglycemia (9).

Laboratory Tests: Additional Procedures and Tests

Although inappropriate elevation of concentration of plasma insulin in the presence of hypoglycemia after overnight or more prolonged fasting remains the basis for the diagnosis of pancreatic islet cell disease, additional tests can confirm the diagnosis when fasting hypoglycemia is mild or absent.

BASAL PROINSULIN Measurement of insulin secretory products (proinsulin, free insulin, and C-peptide) in the peripheral circulation provide additional means of measuring beta cell secretion and prove useful in the study of islet cell function in patients with hypoglycemic disorders (10). Small amounts of proinsulin, the single-chain precursor of insulin, have been found in normal plasma (usually 10–15% and always less than 22% of total immunoreactive insulin). Insulin and connecting peptide (C-peptide) generated from cleavage of proinsulin are secreted in equimolar concentrations. As compared to normal pancreatic tissue, insulin-producing neoplastic tissue typically releases excessive amounts of proinsulin into the blood. The estimation of fasting plasma levels of proinsulin and “proinsulin-like components” may be helpful therefore in diagnosis of islet cell tumors. In 85% of our patients with islet cell tumors, the proinsulin component was clearly elevated and exceeded 25% of fasting total IRI (11). A raised proinsulin concentration is a particularly helpful finding in establishing the diagnosis of insulinoma in patients who have low plasma concentrations of immunoreactive insulin (11, 12). There is insufficient evidence to indicate whether proinsulin is elevated in patients with hyperplasia or nesidioblastosis. In patients with undifferentiated tumors, particularly with islet cell carcinoma, very high proportions of proinsulin are found in plasma.

SUPPRESSION TESTS Suppression of endogenous secretion of insulin, C-peptide, and proinsulin by hypoglycemia induced by exogenous insulin has been used to differentiate between insulin-producing tissue under normal physiological control and autonomous insulin-producing tissue. The test is performed in patients who do not have diagnostic fasting hypoglycemia. When compared with normal subjects, most patients with islet cell tumors show diminished or absent insulin, C-peptide, or proinsulin suppression.

Fish insulin infusion When hypoglycemia was induced by the infusion of fish insulin, suppression of the level of endogenous immunoreactive insulin (measured with an antiserum that also detects proinsulin but does not recognize the administered fish insulin) gave excellent discrimination between patients with insulinomas and normal subjects. Unfortunately, fish insulin is not readily available in the United States (13).

Insulin infusion with C-peptide suppression The infusion of crystalline beef or pork insulin with simultaneous measurements of plasma levels of C-peptide at frequent intervals caused suppression of C-peptide in healthy control subjects but not in the majority of patients with insulin-secreting pancreatic islet cell disease. Probably because there is a long half-life of C-peptide in, as well as biological variability among, normal control subjects, this test is not 100% reliable. In patients with proven islet cell adenoma, normal or near normal suppression of C-peptide has been reported (10, 13). The most useful application of the C-peptide assay has been in situations in which the measurement of insulin is interfered with by the presence of circulating insulin antibodies from administration of insulin in complicating diabetes (10).

Insulin infusion with proinsulin suppression Measurement of plasma human proinsulin during induced hypoglycemia gives excellent discrimination between patients with insulinomas and control subjects (13).

Somatostatin infusion In normal subjects, the administration of somatostatin inhibits tolbutamide-induced insulin release. In some patients with insulinoma, it fails to do so (14), whereas in a patient with hyperinsulinism due to pancreatic beta cell hyperplasia, somatostatin did inhibit tolbutamide-induced insulin release and hypoglycemia. It was thought that this test might differentiate between normal subjects and insulinoma patients (14) and between insulinoma and hyperplasia. In contrast, other authors have found that, in patients with beta cell adenomas, somatostatin completely inhibited tolbutamide-induced release of insulin, but it was incapable of inhibiting the insulin response to glucose in a case of beta cell hyperplasia. Administration of somatostatin does not appear to be reliable in the differential diagnosis of hyperinsulinism.

Adrenalin suppression test Suppression of basal insulin secretion by adrenalin cannot be used as a diagnostic test for insulinoma. Further, such tests are unlikely to distinguish reliably between benign and malignant tumors even though poorly differentiated insulinomas may give lesser suppression with adrenalin (as well as with diazoxide or somatostatin).

PROVOCATIVE TESTS In general, provocative tests are less reliable in the diagnosis of islet cell tumors than are determination of the ratio of immunoreactive insulin to plasma glucose, the suppression tests, or determination of basal proinsulin concentration.

Tolbutamide, leucine, glucagon tests Provocative tests of insulin secretion in patients with suspected islet cell disease are rarely necessary in the differential diagnosis of fasting hypoglycemia. They may be useful, if positive, in some rare insulinoma patients who do not have hypoglycemia after an overnight fast, who can tolerate prolonged fasts without symptoms, and whose levels of serum insulin are not consistently elevated. Examples have been given in recent publications (1, 2). By the plasma insulin criteria described, 80% of patients with proven islet cell disease had abnormal tolbutamide tests, 74% had abnormal leucine tests, and 58% abnormal glucagon tests (1). Since all these tests may yield false negative responses, they are rarely employed today for diagnostic purposes. Excellent differentiation with the tolbutamide test (3) and the glucagon test (15) have been reported by others.

Glucose tolerance test The oral glucose tolerance test is indispensable for the differentiation of the common functional reactive (postabsorptive) hypoglycemias, but usually is of no use in the diagnosis or differential diagnosis of fasting hypoglycemias. However, during a five-hour glucose tolerance test in a patient suspected of having functional hypoglycemia, reactive hypoglycemia not followed by rebound elevation of plasma glucose into the normal range should suggest that excessive insulin was released from autonomous insulin-producing tissue. An example has been given (2).

Calcium infusion The stimulation of insulin secretion by calcium infusion in patients with insulinomas has been suggested as another provocative test for insulinomas. Because of inconsistent results with calcium infusion its value in the diagnosis of insulin-secreting islet cell tumors of the pancreas is questioned.

TESTS FOR MALIGNANCY In patients with malignant insulinoma, increased plasma levels of hCG- α (8 of 14, or 57%), hCG- β (21%), or immunoreactive hCG (25%) were found. Plasma levels of hCG and its subunits were not elevated in any of 41 benign insulinomas (16). Thus an elevated level of hCG or its subunits in a patient diagnosed as having insulinoma will indicate preoperatively the presence of malignant insulinoma.

PLASMA GLUCAGON Some patients with pancreatic islet cell disease secrete not only excessive amounts of insulin but also of glucagon or gastrin (or both) as demonstrated by elevated plasma levels of these hormones (17). The likelihood of such an occurrence is greater in patients with islet microadenomatosis or carcinoma and especially in patients who are members of families with multiple endocrine adenomatosis Type I.

HUMAN PANCREATIC POLYPEPTIDE Insulinomas usually are relatively homogeneous betacytomas and as such do not secrete hormones other than insulin. Plasma levels of human pancreatic polypeptide (hPP) may be elevated in some patients with insulinoma either because there are hyperplastic extratumor pancreatic PP cells or because the tumor may be of mixed (beta and PP cells) cell type. Plasma hPP is elevated in about 50% of patients with insulin-secreting adenomas and does not decrease to normal levels after excision of the tumor (18).

MULTIPLE ENDOCRINE NEOPLASIA TYPE I Insulinoma, benign or malignant, is a feature of multiple endocrine adenomatosis Type I (MEA-I). Therefore in the work-up of patients with insulinoma, tests to exclude parathyroid, pituitary, and adrenocortical hyperfunction are appropriate. Sometimes elevation of hPP in plasma is an indicator of MEA-I in family members otherwise evidently uninvolved by the syndrome.

NONSUPPRESSIBLE INSULIN-LIKE ACTIVITY Severe fasting hypoglycemia has been reported in more than 200 patients harboring nonpancreatic tumors of mesodermal, epidermal, or endodermal origin. High levels of serum immunoreactive insulin have been reported in only two patients who had either a large fibrosarcoma or a bronchial carcinoid with metastases. Some such nonpancreatic tumors evidently synthesize a polypeptide or polypeptides not recognized immunologically as insulin but possessing insulin-like biological activity. Plasma NSILA-S (nonsuppressible insulin-like activity, soluble in acid alcohol) measured by radioreceptor assay has been found to be elevated in one third of such patients (19).

TESTS TO RULE OUT FACTITIOUS HYPOGLYCEMIA The association of an increase in plasma immunoreactive insulin with a low level of plasma glucose is diagnostic of insulin-mediated hypoglycemia. Except in the neonate, this association indicates the presence of pancreatic islet cell disease provided that the surreptitious self-administration of insulin or sulfonyleureas has been ruled out.

Insulin antibodies If insulin has been administered for several weeks or more, circulating antibodies to insulin can usually be demonstrated. When insulin has been administered for short terms (one day to 5 weeks) antibody may not be detectable in the face of severe hypoglycemia due to surreptitious administration of insulin.

C-peptide The radioimmunoassay of serum C-peptide has facilitated the detection of factitious hypoglycemia. The demonstration simultaneously of low plasma glucose, high immunoreactive insulin levels, and suppressed C-peptide immunoreactivity is characteristic of factitious hypoglycemia or hyperinsulinism due to administration of insulin. In endogenous hyperinsulinism, on the other hand, insulin and C-peptide are secreted and therefore plasma levels of both are elevated. Determination of plasma free C-peptide and free insulin permits patients with high titers of insulin antibodies, including those with history of insulin-treated diabetes, to be studied and diagnosed in a way similar to that in subjects who have no circulating insulin antibodies (20).

Plasma sulfonylurea and urinary excretion product Sulfonylurea-induced factitious hypoglycemia can masquerade as an insulinoma both clinically and biochemically, and its incidence may be increasing (21). Hypoglycemia may be accompanied by elevated serum levels of immunoreactive insulin. The similarity between sulfonylurea-induced hypoglycemia and the insulinoma syndrome emphasizes the importance of screening the blood and urine for sulfonylurea agents prior to diagnosis of insulinoma and abdominal exploration. A white precipitate, carboxytolbutamide, which appears in urine when it is acidified, points to the ingestion of tolbutamide.

TESTS FOR INSULIN AUTOIMMUNE SYNDROME Some patients with fasting or severe postprandial hypoglycemia have insulin-binding antibodies in serum but no history of having received exogenous insulin. It has been suggested that these findings are a manifestation of an insulin autoimmune syndrome. During hypoglycemia, large amounts of immunoreactive insulin are extractable from the serum. Hypoglycemia may occur either when saturation of insulin-binding capacity is approached but release of insulin from beta cells continues, such as after a meal, or whenever excessive insulin becomes dissociated from the antigen-antibody complex. It may be possible to distinguish insulin autoimmune syndrome from factitious hypoglycemia due to surreptitious administration of insulin by measuring the concentration of connecting-peptide in plasma. A large amount of C-peptide immunoreactivity (due to proinsulin) has been demonstrated in one patient and indicates that the patient's beta cells were hyperresponsive rather than

suppressed as by exogenous insulin taken surreptitiously (22). This syndrome is not related to the occasional hypoglycemia seen in patients with antiinsulin receptor disorders. Such patients usually have a history of insulin-resistant diabetes and acanthosis nigricans.

LOCALIZATION OF INSULINOMA

In recent years the variety of procedures for the localization of insulinomas and their usage has increased. The results of localization procedures are an aid in planning and performing the surgical treatment; the diagnosis is based upon clinical and biochemical data.

Pancreatic Arteriography

Reports have appeared in which arteriography correctly identified an insulinoma in 38, 54, 71, 73, 80, 83, 86% of the patients. Edis et al (23) reported that in 30 of 34 patients insulin-secreting islet tumors were accurately localized before operation by arteriography with the adjuncts of magnification, stereoscopy, and subtraction techniques. There were no complications of the examination. This percentage was obtained in patients in whom a clinical and biochemical diagnosis of insulinoma had been made earlier. Removal of a tumor localized by arteriography does not necessarily mean however that all insulinomas have been excised; in approximately 10% of cases there is more than one tumor. Nor does arteriography permit exclusion of coexisting adenomatosis or hyperplasia of islets. In addition, in bona fide cases of insulinoma the localization indicated by arteriography is occasionally false.

Portal and Splenic Vein Catheterization and Venous Sampling for IRI

The deficiencies in the arteriographic method of localizing insulinomas have led investigators in recent years to develop methods in which the localization is inferred from the demonstration of much higher concentrations of insulin in veins draining the tumor than in veins draining uninvolved pancreas. The sampling of pancreatic venous blood and its assay for insulin has permitted the localization of small insulinomas or areas of islet hyperplasia that were undetectable by arteriography or by palpation of the pancreas (24, 25).

In one method localization is accomplished by preoperative percutaneous transhepatic catheterization of the portal vein and its pancreatic radicles (24, 25). In another method the splenic (and portal) vein is catheterized and plasma insulin is measured during laparotomy (25). Although experience with these methods is limited, the retrograde catheterization of the portal

vein afforded by the transhepatic route appears to permit a more precise localization, especially of tumors in the head of the pancreas. In this method blood can be obtained from the smaller veins draining the pancreatic head, body, and tail and the position of the catheter confirmed by injection of contrast medium. For practical purposes the intraoperative antegrade catheterization of the splenic vein allows only the splenic and portal vein blood to be sampled; localization of the catheter tip is determined by observing centimeter marks on the catheter. Further, for this method to aid during the course of the operation it is necessary that a rapid insulin assay be available and that the operation be prolonged to accommodate the hour or so needed for the assay to be done. On the other hand the transhepatic method carries the risks attendant upon percutaneous puncture of the liver, e.g. intrahepatic hematoma (24). Localization by either technique can make possible the resection only of the tumor or the affected portion of the pancreas.

Other Localizing Procedures

EMI body scanning (computerized axial tomography) failed to reveal a tumor localized by venous sampling technique (25). EMI scanning, ultrasound, and administration of ^{131}I -toluidine blue-O followed by external scanning have not had sufficient use to evaluate their efficacy as localizing procedures, but it seems unlikely that they will prove more accurate than venous sampling methods. Occasionally duodenography or endoscopic retrograde pancreatography can provide localizing information for appropriately located and sufficiently large tumors.

Liver Scanning

External scanning of the liver following the intravenous administration of technetium sulfur colloid can detect evidence (cold areas) of hepatic metastasis of malignant insulinomas. Such findings help confirm the diagnosis of malignancy and provide objective assessment of the effects of anticancer chemotherapy.

TREATMENT

The treatment of insulinoma for the most part is surgical; in most cases it provides complete cure of insulinoma. It should be performed only when the diagnosis is certain and only by surgeons skilled in pancreatic surgery. Sometimes the preoperative localization of an insulinoma in the head of the pancreas and/or the medical condition of the patient may increase the hazard of surgery sufficiently to justify medical management whenever it has proven to be efficacious. Patients for whom an operation is planned should first have had a short trial of treatment with diazoxide and a diuretic

benzothiadiazine. A successful preoperative trial allows the surgeon to be more conservative during laparotomy should there be difficulty in the identification, location, or removal of the pathologic tissue or should other complications arise. Medical treatment is required for the great majority of malignant insulinomas, since only occasionally will they be cured by operation. Medical treatment for insulinoma is antihormonal therapy and for malignant insulinoma is antihormonal and antitumor therapy.

Medical Management of Benign Tumors

DIET The cornerstone of the medical management of insulinoma is the diet. Not uncommonly patients may avoid symptoms of hypoglycemia for some time by spacing feedings so as to shorten the number of hours between feedings. For some, the inclusion of a bedtime (11 PM) feeding is sufficient; for others a midmorning, midafternoon, or possibly a 3-AM snack will be necessary. Even though the tumor may be stimulated to secrete insulin by the ingestion of carbohydrates, it usually is inadvisable to preferentially restrict the intake of carbohydrate. More slowly absorbable forms of carbohydrate (bread, potatoes, etc), however, are generally preferred. During hypoglycemic episodes rapidly absorbed forms (fruit juices with added glucose or sucrose) are indicated.

DIAZOXIDE AND NATRIURETIC BENZOTHIADIAZINES Diazoxide, now marketed as Proglycem® has proved to be a major advance in the medical management of insulinomas, both benign and malignant (26). It owes its potent hyperglycemic properties to two effects. It directly inhibits the release of insulin by beta cells, and it has an extrapancreatic hyperglycemic effect. Because diazoxide induces the retention of sodium, edema is troublesome at higher dosages. The addition of a diuretic benzothiadiazide, e.g. trichlormethiazide, not only corrects or prevents edema but synergizes the hyperglycemic effects of diazoxide. At the doses needed to counteract higher doses of diazoxide (e.g. 600–1000 mg/day) trichlormethiazide frequently induces hypokalemia. Nausea is an additional complication of higher dosages of diazoxide, and hirsutism may complicate long-term treatment. Nevertheless, patients with benign insulinoma have been successfully managed for up to 12 or more years with diazoxide in doses of 150–600 mg/day in combination with trichlormethiazide in doses of 2–8 mg/day. If they can be tolerated, higher doses may be used in malignant insulinomas.

DILANTIN The anticonvulsant diphenylhydantoin has been shown to inhibit the *in vitro* release of insulin from both the labile and storage beta cell pools. It has been used successfully to control refractory hypoglycemia

as evidenced by normal overnight fasting glucose levels and absence of hypoglycemia during total fasting for up to 24 hours. However in only one third or less of patients with benign insulinoma is the hyperglycemic effect of Dilantin® of any clinical significance. Maintenance doses of Dilantin range from 300 to 600 mg/day. The concurrent administration of diazoxide lowers measurable blood levels of Dilantin and their concurrent use is not recommended.

PROPRANOLOL Only a few reports of the use of propranolol have appeared. Its use has been associated with a reduction in plasma insulin levels and with relief of hypoglycemic attacks in a patient with a benign insulinoma (80 mg propranolol orally per day) and in a patient with malignant insulinoma (640 mg orally per day) in whom streptozotocin was no longer effective.

CHLORPROMAZINE A dosage of chlorpromazine as high as 500 mg per day was associated with decrease in plasma IRI and increase in plasma glucose to euglycemic range during two months' treatment of a patient with malignant insulinoma.

Medical Management of Malignant Tumors

In general the intensities of application of the medical modalities discussed above must be increased when they are applied to the treatment of malignant insulinoma. In addition, sooner or later it becomes necessary to use tumorcidal drugs.

STREPTOZOTOCIN Streptozotocin is the most effective antitumor agent to date for treating malignant insulinomas. It is a broad spectrum antibiotic and a naturally occurring nitrosourea that causes selective destruction of the pancreatic beta cell. It is capable of controlling hypoglycemia but also of causing measurable decrease in tumor size. In 30 cases of functioning malignant insulinoma treated with streptozotocin 50% of patients had an objective reduction in tumor mass, and 17% were considered to have entered complete remission (27). In 39 cases, 64% showed favorable biochemical responses, and in 26%, levels of glucose and insulin returned to normal. Overall, treatment significantly prolonged survival. Unfortunately the specificity of streptozotocin for the pancreatic beta cell is relative; toxicity affects kidney, liver, and the hematopoietic system. Renal tubular toxicity is the most serious and limits the amount of drug that can be administered. It can be avoided usually by monitoring the urine for protein. The next dose should be delayed until there is clearing of proteinuria. Currently streptozotocin is limited to investigational use only in patients

with malignant disease. A weekly schedule of 1–2 g/m² intravenously administered has been used frequently (27).

L-ASPARAGINASE Due to its hepatic and pancreatic toxicity, use of L-asparaginase is limited to the short-term palliation of severe hypoglycemia. It has not been shown to affect tumor mass (28).

5-FLUOROURACIL The experience with 5-fluorouracil is limited to reports of therapeutic effectiveness in several case reports.

TUBERCIDIN Tubercidin has induced regression of tumor mass in several cases of islet carcinoma, but its severe toxicity to veins and tissue surrounding the site of administration limits its usefulness. Renal damage, leukopenia, and thrombocytopenia have also been reported (29).

ADRIAMYCIN Therapy with adriamycin led to significant decreases in insulin levels or in measurable disease; two patients had clinical remissions lasting 5–7 months (30). Cardiac toxicity and reversible myelosuppression are the main toxic effects that limit the dosage of adriamycin, but cumulative doses of less than 500 mg/m² lead to congestive cardiomyopathy in only 0.3% of patients treated with adriamycin.

MITHRAMYCIN A patient unresponsive to regimens of streptozotocin, 5-fluorouracil, adriamycin, tubercidin, phenytoin, and diazoxide was responsive to mithramycin. Plasma insulin sharply declined after 1 mg was given intravenously. Four additional 1-mg doses were followed by normoglycemia for 2 weeks, and no symptoms of hypoglycemia for another 8 weeks before the patient died of apparent congestive heart failure (31).

OTHER MEDICAL MODALITIES The use of continuous intravenous infusion of glucose, coupled with increased dietary intake of carbohydrate, frequently will alleviate hypoglycemia long enough to institute additional therapy. The use of glucocorticoids, which increase gluconeogenesis and insulin resistance, can also help stabilize blood glucose at an acceptable level. Glucagon may help raise blood glucose concentrations but may simultaneously directly stimulate the release of insulin.

CONCLUSION

The diagnostic challenges of today and the future lie in the preoperative differentiation between the various pathological entities that produce the "insulinoma syndrome." Medical treatment modalities must be developed

that are more effective and permit more focused treatments of the benign and malignant forms of organic hyperinsulinism. The advances of the past decade give reason to expect that in the future these diagnostic and therapeutic challenges will be met successfully.

Literature Cited

1. Fajans, S. S., Floyd, J. C. Jr., Vij, S. K. 1975. In *Endocrinology and Diabetes*, ed. L. Kryston, R. Shaw, Ch. 44, pp. 453-72. New York: Grune & Stratton
2. Fajans, S. S., Floyd, J. C. Jr. 1976. Fasting hypoglycemia in adults. *N. Engl. J. Med.* 294:766-72
3. Service, F. J., Dale, A. J. D., Elveback, L. R., Tiang, N.-S. 1976. Insulinoma: Clinical and diagnostic features in 60 consecutive cases. *Mayo Clin. Proc.* 51:417-29
4. Harrison, T. S., Child, C. G., Fry, W. J., Floyd, J. C. Jr., Fajans, S. S. 1973. Current surgical management of functioning islet cell tumors of the pancreas. *Ann. Surg.* 178:485-95
5. Merimee, T. J., Tyson, J. E. 1974. Stabilization of plasma glucose during fasting. Normal variations in two separate studies. *N. Engl. J. Med.* 291:1275-78
6. Felig, P., Lynch, V. 1970. Starvation in human pregnancy: Hypoglycemia, hypoinsulinemia and hyperketonemia. *Science* 170:990-92
7. Merimee, T. J., Tyson, J. E. 1977. Hypoglycemia in man. Pathologic and physiologic variants. *Diabetes* 26:161-65
8. Turner, R. C., Oakley, N. W., Nabarro, J. D. N. 1971. Control of basal insulin secretion, with special reference to the diagnosis of insulinomas. *Br. Med. J.* 2:132-35
9. Bar, R. S., Gorden, P., Roth, J., Sieberg, C. W. 1977. Insulin receptors in patients with insulinomas: Changes in receptor affinity and concentration. *J. Clin. Endocrinol. Metab.* 44:1210-13
10. Rubenstein, A. H., Kuzuya, H., Horwitz, D. L. 1977. Clinical significance of circulating C-peptide in diabetes mellitus and hypoglycemic disorders. *Arch. Intern. Med.* 137:625-32
11. Sherman, B. N., Pek, S., Fajans, S. S., Floyd, J. C. Jr., Conn, J. W. 1972. Plasma proinsulin in patients with functioning pancreatic islet cell tumors. *J. Clin. Endocrinol. Metab.* 35:271-80
12. Alsever, R. N., Roberts, J. P., Gerber, J. G., Mako, M. E., Rubenstein, A. H. 1975. Insulinoma with low circulating insulin levels: The diagnostic value of proinsulin measurements. *Ann. Intern. Med.* 82:347-50
13. Turner, R. C., Heding, L. G. 1977. Plasma proinsulin, C-peptide and insulin in diagnostic suppression tests for insulinomas. *Diabetologia* 13:571-77
14. Lorenzi, M., Gerich, J. H., Karam, J. H., Forsham, P. H. 1975. Failure of somatostatin to inhibit tolbutamide-induced insulin secretion in patients with insulinomas: A possible diagnostic tool. *J. Clin. Endocrinol. Metab.* 40:1121-24
15. Oneda, A., Maruhama, Y., Itabashi, H., Horigone, K., Yanbe, A. 1975. Diagnostic value of intravenous glucagon test in insulinoma. *Tohoku J. Exp. Med.* 116:205-11
16. Kahn, C. R., Rosen, S. W., Weintraub, B. D., Fajans, S. S., Gorden, P. 1977. Ectopic production of chorionic gonadotropin and its subunits by islet-cell tumors. *N. Engl. J. Med.* 297:565-69
17. Hayashi, M., Floyd, J. C. Jr., Pek, S., Fajans, S. S. 1977. Insulin, proinsulin, glucagon and gastrin in pancreatic tumors and in plasma of patients with organic hyperinsulinism. *J. Clin. Endocrinol. Metab.* 44:681-94
18. Floyd, J. C. Jr., Fajans, S. S., Pek, S., Chance, R. E. 1977. A newly recognized pancreatic polypeptide; plasma levels in health and disease. *Rec. Prog. Horm. Res.* 33:519-70
19. Hyodo, T., Megyesi, K., Kahn, C. R., McLean, J. P., Friesen, H. G. 1977. Adrenocortical carcinoma and hypoglycemia: Evidence of production of nonsuppressible insulin-like activity by the tumor. *J. Clin. Endocrinol. Metab.* 44:1175-84
20. Scarlett, J. A., Mako, M. E., Rubenstein, A. H., Blix, P. M., Goldman, J., Horwitz, D. L., Tager, H., Jaspan, J. B., Stjernholm, M. R., Olefsky, J. M. 1977. Fictitious hypoglycemia: Diagnosis by measurement of serum C-peptide immunoreactivity and insulin-binding antibodies. *N. Engl. J. Med.* 297:1029-32
21. Jordan, R. M., Kammer, H., Riddle, M. R. 1977. Sulfonyleurea-induced facti-

- tious hypoglycemia. *Arch. Intern. Med.* 137:390-93
22. Ichihara, K., Shima, K., Saito, Y., Nonaka, K., Tarui, S., Nishikawa, M. 1977. Mechanism of hypoglycemia observed in a patient with insulin autoimmune syndrome. *Diabetes* 26:500-6
 23. Edis, A. J., McIlrath, D. C., Van Heerden, J. A., Fulton, R. E., Sheedy, P. F., Service, F. J., Dale, A. J. 1976. Insulinoma—current diagnosis and surgical management. *Curr. Probl. Surg.* 13:3-45
 24. Ingemansson, S., Köhl, C., Larsson, L.-I., Lunderquist, A., Lundquist, I. 1978. Localization of insulinomas and islet cell hyperplasia by pancreatic vein catheterization and insulin assay. *Surg. Gynecol. Obstet.* 146:725-34
 25. Turner, R. C., Morris, P. J., Lee, E. C. G., Harris, E. A. 1978. Localization of insulinomas. *Lancet* 1:515-18
 26. Fajans, S. S., Floyd, J. C. Jr., Thiffault, C. A., Knopf, R. F., Harrison, T. S., Conn, J. W. 1968. Further studies on diazoxide suppression of insulin release from abnormal and normal islet tissue in man. *Ann. NY Acad. Sci.* 150:261-80
 27. Broder, L. E., Carter, S. K. 1974. Chemotherapy of malignant insulinomas with streptozotocin. *Proc. 8th Congr. Int. Diabetes Fed. Excerpta Med. Int. Congr. Ser.* 314:714-27
 28. Schein, P. S. 1973. Islet cell tumors. Current concepts and management. *Ann. Intern. Med.* 79:239-57
 29. Schein, P. S. 1972. Chemotherapeutic management of the hormone-secreting endocrine malignancies. *Cancer* 30: 1616-26
 30. Eastman, R. C., Come, S. E., Strewler, G. J., Gorden, P., Kahn, C. R. 1977. Adriamycin therapy for advanced insulinoma. *J. Clin. Endocrinol. Metab.* 44:142-48
 31. Kiang, D. T., Frenning, D. H., Bauer, G. E. 1978. Mithramycin for hypoglycemia in malignant insulinoma. *N. Engl. J. Med.* 299:134-35

VITRECTOMY SURGERY FOR DIABETIC RETINOPATHY

◆7327

George H. Bresnick, M.D. and Franklin L. Myers, M.D.

Department of Ophthalmology, University of Wisconsin,
Madison, Wisconsin 53706

INTRODUCTION

The vitrectomy operation is a new ophthalmic surgical procedure for the treatment of diabetic retinopathy. Since vitrectomy is a major intraocular surgical procedure with the potential for serious ocular complications, its use is reserved for certain advanced stages of proliferative diabetic retinopathy (PDR) where safer alternative modes of therapy (e.g. photocoagulation) are ineffective. With this in mind one may consider vitrectomy surgery for two aspects of severe diabetic retinopathy: (a) severe vitreous hemorrhage, and (b) retinal detachment affecting or threatening central vision.

This review article delineates the current techniques, indications, and results of vitrectomy surgery in the treatment of diabetic retinopathy. The natural history and a description of other modes of treatment of diabetic retinopathy introduce the discussion of the surgical procedure. Since new developments in this area of ophthalmic surgery are occurring rapidly, the potential of future advances is also discussed. The medical problems encountered in caring for the diabetic patient undergoing vitrectomy may be considerable, and therefore this aspect is also reviewed.

NATURAL HISTORY OF DIABETIC RETINOPATHY

Nonproliferative and Proliferative Phases

The fundus abnormalities seen in diabetic retinopathy consist of changes within the retina (intraretinal), in front of the retina (preretinal), and within the vitreous cavity (intravitreal). The intraretinal changes comprise the nonproliferative phase of the disease, while the preretinal and intravitreal alterations constitute the proliferative phase. Many of the nonproliferative

changes are manifestations of retinal ischemia due to the occlusion of capillaries and small arterioles within the retina. (The mechanism of retinal vaso-occlusion is poorly understood.) These occlusive changes precede or coincide with the development of retinal neovascularization; the retinal ischemia that results from small-vessel occlusion is thought to stimulate the growth of new vessels and fibrous tissue during the proliferative phase of the disease (1).

Abnormally increased permeability of the retinal capillaries is another fundamental aspect of the nonproliferative phase of the disease. Retinal edema and hard exudate deposits are the major consequence of the abnormal retinal vascular permeability.

Causes of Visual Loss from Diabetic Retinopathy

Visual loss from the nonproliferative component of the disease may result from macular ischemia or from retinal edema and hard exudates involving the macula. No treatment is available for macular ischemia; retinal photocoagulation directed at abnormally permeable capillaries and microaneurysms may be of benefit, however, in certain cases of macular edema (2, 3).

The most severe visual loss from diabetic retinopathy occurs from the proliferative aspects of the disease. Vitreous hemorrhage and tractional detachment of the retina are the two major causes of this severe visual loss; both result from the growth of fibrovascular tissue and the subsequent shrinkage of this tissue and of the vitreous.

The fibrovascular tissue originates from the retinal vessels, breaks through the internal limiting membrane of the retina, and proliferates at the interface between retina and vitreous. The fibrovascular tissue is intimately connected to the cortical (outer) vitreous layer.

For unknown reasons the vitreous and fibrovascular tissue contract, pulling the newly formed fibrovascular tissue forward. This traction exerted on the new vessels by the contracting vitreous is a major cause of vitreous hemorrhage in PDR (4, 5). Vitreous contraction may also produce traction on the retina itself at sites of adhesion between vitreous, fibrovascular tissue, and retina; tractional detachment of the retina may result. If the retinal detachment involves the macula, severe visual loss may occur.

TREATMENT OF PROLIFERATIVE DIABETIC RETINOPATHY

Photocoagulation Treatment

The first line of treatment of PDR is photocoagulation. The results of a recent collaborative trial of photocoagulation (6) show that this treatment

can significantly reduce the incidence of severe visual loss from PDR by eliminating or reducing neovascularization and by curtailing the proliferative process. Nevertheless, some eyes with PDR develop visual loss from vitreous hemorrhage and/or tractional retinal detachment in spite of photocoagulation, and other eyes are considered untreatable with photocoagulation because they already have severe vitreous hemorrhage or severe vitreoretinal traction. It is these two groups of eyes that are potential candidates for vitrectomy surgery.

Indications for Vitrectomy Surgery

VITREOUS HEMORRHAGE The indications for vitrectomy surgery in eyes with vitreous hemorrhage depend upon the severity and duration of the hemorrhage. Many of these eyes (especially those with mild to moderate degrees of hemorrhage) will clear spontaneously over a period of a few weeks to several months and will not require surgery. For this reason, most ophthalmic surgeons do not perform vitrectomy surgery for vitreous hemorrhage unless the vision is severely reduced ($<5/200$ visual acuity) and, unless the hemorrhage has been present for a least 6–12 months.

TRACTIONAL DETACHMENT OF THE RETINA Tractional detachment of the retina in PDR usually begins as a small area of retinal elevation located at some distance from the macula. In many eyes the detachment will not progress or will progress only slowly, and unless the detachment extends to include the macula, good visual acuity will be maintained. For this reason, surgery is not undertaken for tractional retinal detachment unless the macula is detached or imminently threatened. Should vision decrease because of macular detachment, prompt surgical repair is important to restore useful vision. Conventional scleral-buckling surgery may be used in some cases of retinal detachment in PDR (7), but vitrectomy surgery is required if the degree of vitreoretinal traction is severe.

HISTORICAL DEVELOPMENTS AND TECHNIQUES OF VITREOUS SURGERY

Traditionally, a surgical approach to diseases involving the vitreous body of the eye was considered too dangerous to undertake because one could easily produce retinal tears and retinal detachment by inadvertent traction on the collagenous vitreous framework. A dramatic breakthrough occurred in 1972 when Machemer et al (8) introduced an instrument called the vitreous infusion suction cutter (VISC) that could cut and remove vitreous with a minimum amount of vitreoretinal traction. The features of the instrument were:

1. A small-diameter (2.6 mm) tube with a motor-driven rotating internal cutting device that could safely remove small beads of vitreous.
2. An infusion system that replaced the vitreous with an equivalent amount of normal saline (or other biologically compatible fluid) such that a reasonably constant volume and pressure were maintained in the eye.
3. An illuminated operating microscope and corneal contact lens combination that allowed direct visualization of the vitreous and retina during the procedure.

The instrument was introduced into the eye through a small anterior sclerotomy incision located over the pars plana of the ciliary body. The cutting and sucking action of the instrument allowed the removal of blood and vitreous from the vitreous cavity and the severance of vitreoretinal traction bands. In this way the operation could perform two basic functions: (a) removal of vitreous opacities (e.g. hemorrhage) and (b) release of vitreoretinal traction to allow settling of tractional retinal detachment.

Since the introduction of the VISC a number of similar devices have been designed (9-13) with modifications that attempt to (a) reduce the size of the instrument, (b) improve the cutting capability of the instrument, (c) provide an intraocular fiber optics illumination system, and (d) provide intraocular diathermy and cryocoagulation.

These and similar refinements have extended the capabilities of the procedure and reduced some of the complications. Delicate surgical maneuvers such as peeling of opaque membranes from the surface of the retina and intraocular coagulation of intravitreal new vessels are now feasible.

RESULTS OF VITRECTOMY SURGERY

Early in the development of vitreous surgery most of the cases operated were for long-standing (one year or longer) severe vitreous hemorrhage. Reports indicated that in 60-70% of these cases the visual acuity could be improved by surgery (14, 15). However, the level of visual acuity achieved was often disappointing, with less than 3% recovering 20/40 or better vision (16). Retinal degenerative changes secondary to ischemia, and macular detachment or fibrosis, accounted for most of the more severe visual loss.

As the backlog of cases with long-standing vitreous hemorrhage was gradually reduced, surgeons, hoping to improve the visual outcome, began operating earlier in the course of vitreous hemorrhage. Results of earlier surgery are currently under study (see below). Attention was also turned to surgery in so-called clear media cases in which relief of tractional retinal detachment that involved or threatened the macula was the primary aim. Results in these cases are more difficult to assess, since success is judged

partly by the anatomical result as well as by the visual result. The outcome in individual cases can sometimes be excellent, but the overall success rate appears to be no greater than 40–50% (17–19). Nevertheless, when confronted with the inevitable blindness that results if reattachment of the retina is not attempted, most surgeons and patients are willing to proceed with surgery even if the odds for success appear relatively low.

COMPLICATIONS OF VITRECTOMY SURGERY

In addition to being influenced by the underlying microvascular and vitreous pathology present (degree of capillary closure, extent and sites of vitreoretinal adhesions, status of the macula, etc), the success rate is also considerably influenced by the complication rate. Complications include iatrogenic retinal detachment, corneal dystrophy, cataract, recurrent vitreous hemorrhage, and rubeosis iridis with neovascular glaucoma.

Retinal detachment that develops during surgery is usually caused by inadvertent tearing of the retina by the vitrectomy instrument. While repair of the detachment is possible in some cases, large tears may produce irreparable total retinal detachments. Improved visualization with fiber optic attachments and improved cutting capabilities of newer instruments have reduced this complication.

Corneal dystrophy due to corneal endothelial damage during surgery is most common when lensectomy (cataract removal) and vitrectomy are combined. Reduced intraocular operating time and the introduction of newer intraocular irrigating solutions seem to have decreased this complication (20).

Cataract may develop during surgery when the lens is inadvertently hit with the vitrectomy instrument or when a large quantity of irrigating solution has been used. The lens opacities in the latter instance are usually reversible. However, if lens changes develop during surgery sufficient to impair visualization of the fundus, the lens may have to be removed to allow safe completion of the case.

Recurrent vitreous hemorrhage following vitrectomy may result from residual retinal neovascularization or from neovascular ingrowth at the sclerotomy site. Fortunately, most of these hemorrhages will clear spontaneously, but occasionally a repeat vitreous washout is required.

Ten to twenty percent of diabetic eyes undergoing vitrectomy may develop neovascular glaucoma, usually within six months of surgery. Medical and surgical therapy for neovascular glaucoma is usually unsuccessful; most eyes lose useful vision, and some eventually must be removed for intractable pain. Vitrectomy, therefore is a risky procedure in terms of ocular complications. These obviously must be taken into consideration and balanced against the potential benefits before undertaking surgery.

MEDICAL PROBLEMS

An additional risk of vitrectomy in the diabetic patient involves the risk of surgery and anesthesia itself. Most cases are done under general anesthesia because the surgery may be lengthy (2–5 hours), and immobilization of the patient after introduction of the vitrectomy instrument into the eye is critical. Nonetheless some surgeons find anesthesia with a retrobulbar block combined with intravenous sedation under continuous anesthetist surveillance an acceptable alternative. The great majority of these patients are suffering from the other small- and large-vessel complications of diabetes mellitus, which involve the renal, cardiac, and peripheral vascular systems. Hypertension and cardiac abnormalities are especially pertinent to anesthetic risk and must be carefully evaluated and controlled preoperatively. Renal disease, while certainly not a contraindication to surgery and anesthesia, must be evaluated, and if required, dialysis instituted so that the patient is in the best possible condition for surgery.

Chronic infection, poor understanding by the patient of his disease, and psychological factors all contribute to poor diabetic control that so many of these patients manifest. Although tight control preoperatively is not absolutely necessary, knowledge of the patient's insulin requirements, electrolyte situation, and previous responses to treatment is required in order to smooth out the postoperative management in a period of time when the patient is inactive, may be vomiting from the effects of anesthesia, or may be in pain and not eating.

It is a good practice to admit the patient to the hospital three or four days preoperatively for medical assessment, especially if he comes from a long distance to a medical center and is not being actively managed as an outpatient by the medical consultant. Collection of a 24-hour urine specimen starting the day before admission and completed on the day of admission is helpful in determining current control.

The practice of giving the patient one half to two thirds of his usual long-acting insulin dosage on the morning of surgery along with intravenous glucose seems to work well. Scheduling the surgery as early in the day as possible is helpful, and monitoring of the blood sugar at intervals during surgery is essential.

Fortunately most of these patients have little pain postoperatively, can be ambulated rapidly, and can be returned to their preoperative, long-acting insulin dosage the day after surgery. Occasionally, extensive scleral-buckling surgery for retinal detachment must be done, which will result in more pain, postoperative nausea and vomiting, and a more stormy and protracted course. The so-called rainbow-insulin schedule based on urine sugar is best avoided postoperatively; supplemental regular insulin dosage should be based on additional blood sugar determinations.

Vitrectomy is an elective procedure, and although the patient may be anxious for surgery since it holds out hope for his vision, he must be made aware that management of his whole person takes precedence. Coordination between surgeon, internist, and anesthetist are vital if the rehabilitation of the eye is to be of lasting value to the patient.

FUTURE DEVELOPMENTS

Technical advances in the field have continued since the initial breakthrough by Machemer (8). Smaller instruments, better cutting action, improved visualization, and refined surgical techniques have all been developed but could be improved. The use of lasers for vaporization of the vitreous and intraocular photocoagulation may add to the surgical armamentarium. More important, however, are current efforts to refine the indications of surgery and determine the proper timing of surgery. A cooperative clinical trial, the Diabetic Retinopathy Vitrectomy Study (DRVS), is currently being sponsored by the National Eye Institute. Now entering its third year of patient recruitment, it seeks to determine whether it is better to do immediate vitrectomy on diabetic eyes with severe vitreous hemorrhage (vision $<5/200$) of less than six months' duration, or whether surgery should be deferred until at least 12 months or more following the hemorrhage. While earlier surgery may result in a higher yield of eyes with visual acuity of 20/40 or better, it may also have a higher complication rate because more active retinopathy may be present. Deferring surgery may allow the active process to subside and thus result in a better overall salvage. Since the risks and benefits of either early or late surgery are believed to be in balance, a randomized controlled trial is being done, with eyes being randomized to either early or deferred surgery. Deferred surgery is only done if there is no clearing of the vitreous hemorrhage or if traction detachment of the macula develops. The final visual acuities and the complication rates between the two groups will then be compared. No results are yet available and patient recruitment continues.

In another part of the DRVS, eyes with severe active neovascularization as well as vitreous detachment, and eyes with traction retinal detachment are being observed to determine their natural history. Conventional photocoagulation treatment is applied when indicated. If enough of these eyes progress to severe visual loss so that the deterioration approximates or is worse than the complication rate from surgery, then (in a subsequent study) half of the eyes with this kind of retinopathy will be randomized to vitrectomy while the other half will be followed without vitrectomy unless severe vitreous hemorrhage or macula detachment develops.

This type of cooperative randomized controlled clinical trial is difficult to design and execute; however, it is the most promising approach to obtain

adequate answers to the difficult clinical questions these severely diseased eyes pose. Ultimately, of course, new technology and surgical expertise can only accomplish so much, even when best employed. The final answer lies in the prevention of diabetes and its complications. At the moment that prevention appears to be some time in coming, and vitrectomy remains as a fascinating and challenging "holding action."

Literature Cited

1. Davis, M. D., Myers, F. L., Engerman, R., deVenecia, G., Magli, Y. L. 1969. Clinical observations concerning the pathogenesis diabetic retinopathy. In *Symposium on the Treatment of Diabetic Retinopathy*, ed. M. F. Goldberg, S. L. Fine, Chap. 4. Washington DC, Public Health Serv. Publ. No. 1890
2. Patz, A., Schatz, H., Berkow, J. W., Gittlesohn, A. M., Ticho, U. 1973. Macular edema: An overlooked complication of diabetic retinopathy. *Trans. Am. Acad. Ophthalmol. Otolaryngol.* 77:34
3. Photocoagulation in treatment of diabetic maculopathy: Interim report of a multicentre controlled study. 1975. *Lancet* 2:1110
4. Davis, M. D. 1965. Vitreous contraction in proliferative diabetic retinopathy. *Arch. Ophthalmol.* 74:741
5. Tolentino, F. I., Lee, P. F., Schepens, C. L. 1966. Biomicroscopic study of vitreous cavity in diabetic retinopathy. *Arch. Ophthalmol.* 75:238
6. The Diabetic Retinopathy Study Research Group. 1978. Photocoagulation treatment of proliferative diabetic retinopathy: The second report of diabetic retinopathy study findings. *Am. J. Ophthalmol.* 85:82
7. McMeel, J. W. 1971. Diabetic retinopathy: Fibrotic proliferation and retinal detachment. *Trans. Am. Ophthalmol. Soc.* 69:440
8. Machemer, R., Parel, J. M., Buettner, H. 1972. A new concept for vitreous surgery. 1. Instrumentation. *Am. J. Ophthalmol.* 73:1
9. Douvas, N. G. 1973. The cataract roto-extractor (A preliminary report) *Trans. Am. Acad. Ophthalmol. Otolaryngol.* 77:OP-792
10. Peyman, G. 1973. Experimental vitrectomy, new technical aspects. *Am. J. Ophthalmol.* 75:774
11. Parel, J. M., Machemer, R., Aumayr, W. 1974. A new concept for vitreous surgery. 4. Improvements in instrumentation and illumination. *Am. J. Ophthalmol.* 77:6
12. O'Malley, C., Heintz, R. M. 1975. Vitrectomy with an alternative instrument system. *Am. J. Ophthalmol.* 7:585
13. Charles, S., White, J., Dennison, C. et al. 1976. Bimanual, bipolar intraocular diathermy. *Am. J. Ophthalmol.* 81:101
14. Aaberg, T. M. 1977. Results of 100 consecutive vitrectomy procedures. In *New and Controversial Aspects of Vitreoretinal Surgery*, ed. A. McPherson, pp. 245-49. St. Louis: Mosby
15. Myers, F. L., Bresnick, G. H., Brightbill, F. S. 1977. Vitrectomy in proliferative diabetic retinopathy: An analysis of 60 consecutive tests. See Ref. 14, pp. 326-36
16. Diabetic Retinopathy Vitrectomy Study 1976. Manual of operations, Chapter 2. DRVS Coordinating Center. Minneapolis: Univ. Minn.
17. Mandelcorn, M. S., Blankenship, G. W., Machemer, R. 1976. Pars plana vitrectomy for the management of severe diabetic retinopathy. *Am. J. Ophthalmol.* 81:561-70
18. Michels, R. G. 1978. Vitrectomy for complications of diabetic retinopathy. *Arch. Ophthalmol.* 96:237-46
19. Blankenship, G. W., Machemer, R. 1978. Pars plana vitrectomy for management of severe diabetic retinopathy. *Am. J. Ophthalmol.* 85:553-59
20. Edelhauser, H. F., Van Horn, D. L., Aaberg, T. M. 1976. Intraocular irrigating solutions and their use for vitrectomy. In *Advances in Vitreous Surgery*, ed. A. R. Irvine, C. O'Malley, pp. 265-87. Springfield, Ill: Thomas

DIABETIC KETOACIDOSIS: Reappraisal of Therapeutic Approach

◆7328

*Abbas E. Kitabchi, Ph.D., M.D., Ruth Young, M.D.,
Harold Sacks, M.D., and Lawrence Morris, M.D.*¹

Departments of Medicine and Biochemistry, and Clinical Research Center,
University of Tennessee Center for the Health Sciences, Memphis,
Tennessee 38163

INTRODUCTION

Several comprehensive reviews (1-5) on diabetic ketoacidosis (DKA) have recently appeared in which not only has the much-debated issue of low dose insulin therapy been covered, but also controversial subjects such as pathogenesis of DKA, supportive therapy, and insulin resistance. For detailed discussion of these topics, readers are referred to these reviews. In this review we present an alternate approach in that we (a) review the studies comparing low dose and conventional doses of insulin therapy; (b) discuss the outcome of randomized controlled studies from our Center on various forms of DKA therapy, which further demonstrate the effectiveness of low dose insulin, provide a rational basis for the recommended route of therapy, and elucidate the possible mechanism for occasional failure to respond to therapy; and (c) discuss selected facets of fluid and electrolyte therapy.

After examining reported comparative studies on low, intermediate, and high dose insulin, one clearly finds that the number of well-controlled studies with adequate data base are limited. To our knowledge, there are only six such studies at the time of writing, including one of our own. Only five of these are prospective, while the sixth (8, 9) is retrospective in nature. The results of these data are presented in Tables 1a and 1b. It is clear from

¹This work was supported in part by GRS Grant USPHS RR-05423, General Clinical Research Centers Grant RR00211, and a research grant from Eli Lilly Company, Indianapolis, Indiana.

Table 1a Comparative studies on the efficacy of intermediate and high dose insulin in the treatment of diabetic ketoacidosis

	Smith & Martin 1954 (6)			Shaw et al 1962 (7)	
	Intravenous bolus injections			40% IV and 60% SC	
	Intermediate	Moderate	High	Intermediate	High
Patients (no.)	12	18	13	15	15
Average age (yr)	34.6	48.7	45.3	37.5	41.6
Initial insulin dose (U)	80	160	240	1.5 U/kg 4th hr	200–400/hr
Subsequent insulin dose (U/time)	80/2hr	160/2hr	240/2hr	(variable)	25–100/hr
Serum insulin: Initial (μ U/ml)	—	—	—	—	—
Highest level (μ U/ml)	—	—	—	—	—
Initial plasma glucose (mg/dl)	688	755	820	712	700
Fall in glucose (mg/dl/hr)	—	—	—	91	78
Blood pH	—	—	—	—	—
Serum HCO_3^- (meq/liter)	6.4	5.4	7.0	4.8	5.3
Blood ketones (mM)	—	—	—	15.6	18.4
Time for glucose to reach 250–300 mg/dl (hr)	5.9	5.9	6.1	8.0	8.0
Time for HCO_3^- to reach 15–20 meq/liter (hr)	8.0	8.0	5.2	—	—
Hypoglycemia (%)	—	—	—	0	13
Hypokalemia (%)	8	0	8	—	—
Death (%)	17	0	8	0	13
Comments	Hypokalemia was the major cause of death				

these studies that deficiencies in documentation of data make definitive conclusions impossible. The available results so far suggest that the metabolic responses of DKA to loading bolus plus low dose insulin (5–10 U/hr) are similar to those to higher doses (15–150 U/hr). While the efficacy of a low dose is established, more prospective, controlled studies are warranted to decide the superiority of one mode of treatment over the other. In particular, a plea is made for careful presentation of data, especially the total dosage required for control, incidence of complications, mortality rate, and biochemical criteria for diagnosis and recovery.

The studies concerning insulin dosage and its route of administration from our Center deal with the following five subjects:

1. Low dose intermittent intramuscular (IM) (Protocol 1, Table 2) vs high dose intravenous (IV) and subcutaneous (SC) (Protocol 1, Table 3) insulin in the treatment of adult patients with diabetic ketoacidosis (10)
2. Comparison of the effectiveness of low dose insulin therapy in DKA by three routes of administration (Protocol 2, Table 4) (13)
3. Efficacy of low dose continuous IV infusion of insulin compared with that of intermittent IM administration preceded by a priming bolus of IV and IM insulin (Protocol 3, Table 5) (14)

Table 1b Comparative studies on the efficacy of low dose versus high dose insulin in the treatment of diabetic ketoacidosis

	Alberti et al 1973 (8, 9) ^a		Kitabchi et al 1976 (10)		Heber et al 1977 (11)		Peters et al 1977 (12)			
	Low (IM) ^b	High	Low (IM) ^b	High SC and IV ^b	Low IV infusion	High IV and SC ^b	Very low IV infusion	Low IV infusion	High IV bolus	High IV bolus
Patients (no.)	14	37	24	24	10	7	7	9	10	10
Average age (yr)	47.5	48.2	38.7	40.3	27	40	— ^c	— ^c	— ^c	— ^c
Initial insulin dose (U)	14	77	14	87	6	50-300	0	0	0	50
Subsequent insulin dose (U/time)	5/hr	variable	5/hr	15-50/hr	6/hr	50-300/2hr	2/hr	10/hr	50/2hr	50/2hr
Serum insulin: initial (μ U/ml)	7.1	6.6	23.8	21.2	12.0	—	9	9	9	9
Highest level (μ U/ml)	93	540	89.3	5448	162	—	25	136	66	66
Initial plasma glucose (mg/dl)	637	661	723	697	574	712	671	635	754	754
Fall in glucose (mg/dl/hr)	76	81	71	99	115	114	46	73	78	78
Blood pH	7.22	7.09	7.13	7.15	7.14	7.09	7.19	7.14	7.14	7.14
Serum HCO ₃ ⁻ (meq/liter)	10.2	—	7.7	8.1	9	9	7.4	6.2	5.8	5.8
Blood ketones (mM)	12.1	10.0	14.3	13.7	—	—	16.4	16.5	18.5	18.5
Time for glucose to reach 250-300 mg/dl (hr)	6.5	6.3	6.7	4.5	3.1	3.9	7.0	4.3	3.4	3.4
Time for HCO ₃ ⁻ to reach 15-20 meq/liter (hr)	—	—	11.1	11.6	—	—	>12	8.3	8.9	8.9
Hypoglycemia (%)	0	—	0	25	—	—	—	—	—	—
Hypokalemia (%)	—	—	5	29	—	—	—	—	—	—
Death (%)	7	5	0	0	0	0	—	—	—	—

^a These studies are retrospective.^b Intermittent injection.^c Average age of all patients 37 $\pm$ 3 years.

Table 2 Protocol 1 low dose insulin treatment for DKA patients (from 10)

Initial insulin (IM) ^a (U/kg body weight)	Follow-up insulin (IM) ^b (U/hr)
0.22	5

^aAdministered with a 1.5-inch needle in the deltoid area.^bIf plasma glucose does not fall by 10% of its initial value, repeat the initial bolus of insulin.**Table 3** Protocol 1 high dose insulin treatment for DKA patients (from 10)

Plasma glucose (mg/dl)	Initial insulin (IV + SC) (U)	Follow-up insulin (SC) ^a (U/hr)
> 1000	50 + 100	50
600–1000	25 + 75	50
400– 599	15 + 50	50
300– 399	10 + 30	15

^aIf plasma glucose does not fall by 10% of the initial value in an hour, repeat the initial bolus of insulin.**Table 4** Protocol 2 low dose insulin treatment for DKA patients (13)

Initial insulin (U/kg body weight)	Initial route	Follow-up insulin ^a (U/hr)	Follow-up route
0.33	IM	7	IM injection
0.33	IV	7	IV infusion in saline with albumin
0.33	SC	7	SC injection

^aIf plasma glucose does not fall by 10% of the initial value in an hour, repeat the initial bolus of insulin.**Table 5** Protocol 3 low dose insulin treatment for DKA patients (14)

Initial insulin (U/kg body weight)	Initial route	Follow-up insulin ^a (U/hr)	Follow-up route
0.44	½ IV + ½ IM	7	IM injection
0.44	IV	7	IV infusion in saline with no albumin

^aIf plasma glucose does not fall by 10% of the initial value in an hour, repeat the initial bolus of insulin.

4. Low dose vs high dose infusion of insulin in the treatment of diabetic ketoacidosis in children (15); and
5. Efficacy of low dose insulin therapy of DKA in nonspecialized community hospital under direction of practicing internists (16).

Studies 1, 2, and 3 were designed to provide a randomized protocol for each investigation. The working biochemical definition of DKA in those studies was defined previously (10) and consisted of (a) plasma or serum glucose ≥ 300 mg/dl, (b) plasma $\text{HCO}_3^- < 15$ meq/liter, (c) arterial pH < 7.3 , and (d) serum acetone positive at 1:2 dilution, glycosuria with ketonuria, and age range of 15 years and older. The summary of this protocol is presented in Tables 2–5. Detailed clinical and biochemical profiles of all these patients for each protocol are presented in Table 6. All these values were obtained after patients had received only saline infusion in the emergency room prior to other therapy. As can be seen, there is remarkable consistency in each of the mean values in defined measurements between all the groups.

Figure 1 shows a summary of the responses of plasma glucose, ketone bodies, and glucagon in each group and demonstrates the similarity of the responses in each protocol regardless of the type of therapy. With the exception of a significant lack of the decline in the first two hours for the ketone bodies in the IM and SC groups of Protocol 2, all other parameters were comparable for all groups. This delay was prevented when simultaneous IV and IM injections of insulin as the initial method of therapy (Protocol 3) were replaced for IM or SC priming dose insulin.

Figure 2 provides the detailed recovery parameters of each patient in all protocols as defined by plasma glucose of less than 250 mg/dl, $\text{HCO}_3^- > 15$ meq/liter, and pH > 7.30 . As can be seen, the mean $\pm$ SEM, which are indicated on the top of each appropriate column on Figure 2, are reasonably similar in all groups.

Figure 3 depicts the basal plasma immunoreactive insulin (IRI) and glucose values and the IRI and glucose responses to insulin therapy of patients in Protocols 1 (high and low dose) and 2 (IV, IM, and SC low dose) not previously receiving insulin. Although high dose (SC and IV, or initial IV insulin injections) provided a much higher plasma IRI level than the low dose IM or SC loading doses, the plasma glucose responses in all groups were relatively similar, which indicates low dose IRI levels (60–100 $\mu\text{U/ml}$) adequately achieved the same glucose-lowering effect as the higher doses of insulin (IRI 400–1200 $\mu\text{U/ml}$). In the low IM and SC insulin protocols, however, the IRI values did not reach the plateau level until 2–3 hours after initiation of therapy.

Table 6 Clinical and biochemical profile of DKA patients on admission^a (see Tables 2-5 for detail of protocols)

	Protocol 1		Protocol 2 (low dose)			Protocol 3 (low dose)		
	High dose	Low dose	IM	SC	IV	IV	IV	IM
Number of patients	24	24	15	15	15	15	15	15
Age (Range) (yr)	40.3 (15-70)	38.7 (15-67)	40.7 (19-64)	44.3 (28-75)	37.2 (21-69)	40.6 (20-70)	35.2 (20-68)	
Glucose (mg/dl)	697 ± 59	723 ± 61	523 ± 57	579 ± 62	590 ± 58	548 ± 53	584 ± 52	
Sodium (meq/liter)	137.4 ± 1.7	136.5 ± 1.4	134 ± 2	136 ± 2	138 ± 2	132 ± 2	134 ± 1	
Potassium (meq/liter)	5.4 ± 0.2	5.9 ± 0.3	5.5 ± 0.4	5.8 ± 0.3	6.0 ± 0.3	5.7 ± 0.3	5.9 ± 0.3	
Bicarbonate (meq/liter)	8.1 ± 0.7	7.7 ± 0.5	4.3 ± 0.7	5.7 ± 0.8	5.5 ± 0.8	6.6 ± 1.0	6.1 ± 1.0	
BUN (mg/dl)	30 ± 5	35 ± 4	33 ± 5	28 ± 4	31 ± 4	22 ± 3	27 ± 5	
pH	7.15 ± 0.02	7.13 ± 0.02	7.09 ± 0.02	7.11 ± 0.03	7.10 ± 0.02	7.14 ± 0.05	7.03 ± 0.04	
Cortisol (μg/dl)	72 ± 8	78 ± 6	44 ± 7	58 ± 8	68 ± 8	51 ± 8	75 ± 12	
Glucagon (pg/ml)	409 ± 53	592 ± 90	588 ± 140	543 ± 167	377 ± 74	649 ± 241	903 ± 263	
β-Hydroxybutyrate (mM)	10.3 ± 0.5	11.3 ± 0.6	9.6 ± 0.6	10.0 ± 1.0	10.0 ± 1.0	8.7 ± 1.2	8.8 ± 1.2	
Acetoacetate (mM)	3.4 ± 0.3	3.0 ± 0.2	2.9 ± 0.3	3.7 ± 0.4	3.0 ± 0.3	2.8 ± 0.22	2.6 ± 0.4	
Pyruvate (mg/dl)	0.8 ± 0.1	1.0 ± 0.2	0.89 ± 0.2	0.66 ± 0.07	0.74 ± 0.09	0.5 ± 0.1	0.5 ± 0.1	
Lactate (mg/dl)	20.6 ± 2.2	25.8 ± 2.4	25.4 ± 5.4	22.2 ± 2.4	20.0 ± 2.5	20.1 ± 4.1	23.7 ± 3.2	

^aValues, when applicable, are mean ± SEM after initial hydration in the emergency room prior to any other therapy.

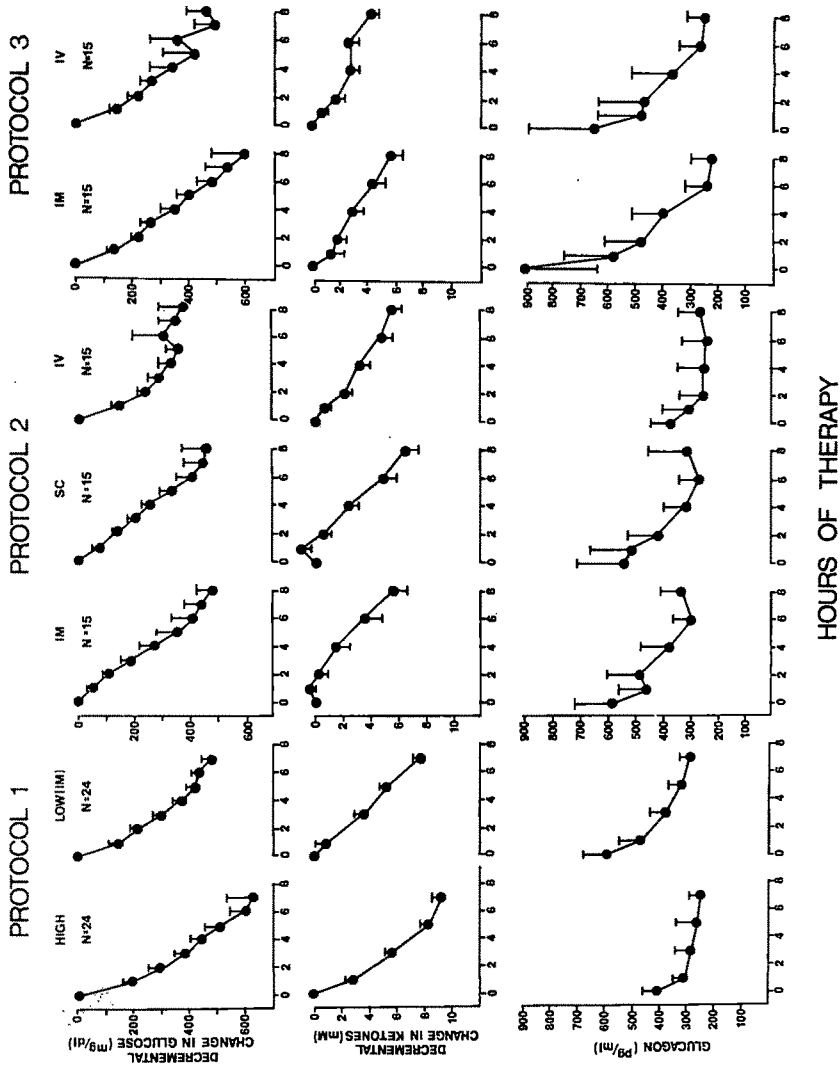


Figure 1 Decremental changes in plasma glucose (mg/dl) and ketone bodies (mM) and fall in plasma glucagon. Conventional method of insulin therapy is compared with six methods of low dose regimens in seven groups of patients (total of 123 patients) from our center from June 1974 thru February 1978. Details of protocol are presented in Table 2. Data from References 10, 13, and 14.

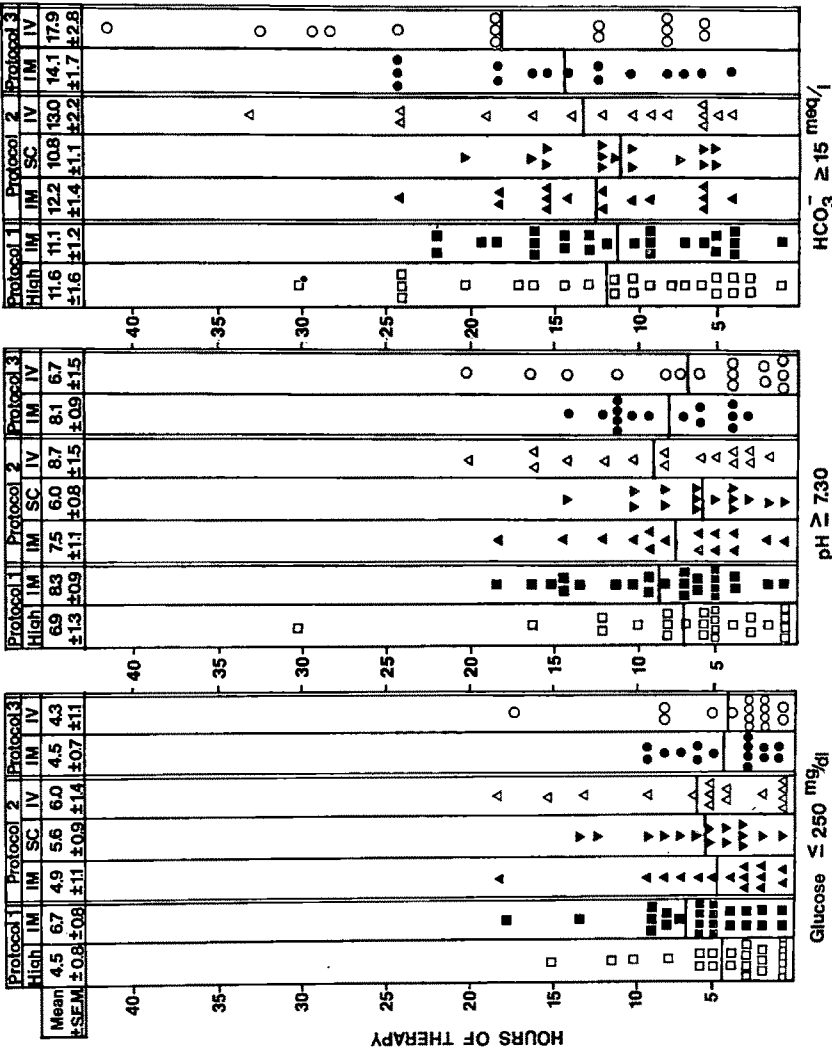


Figure 2 Scattergram of response of each patient to particular method of insulin therapy during diabetic ketoacidosis. The response parameters measured are: hours for plasma glucose to reach 250 mg/dl, serum HCO₃⁻ to reach 15 meq/liter, and blood pH to reach 7.3. Details of each protocol are in Table 2. Mean $\pm$ sem for each group are presented on top of each column. Data from References 10, 13, and 14.

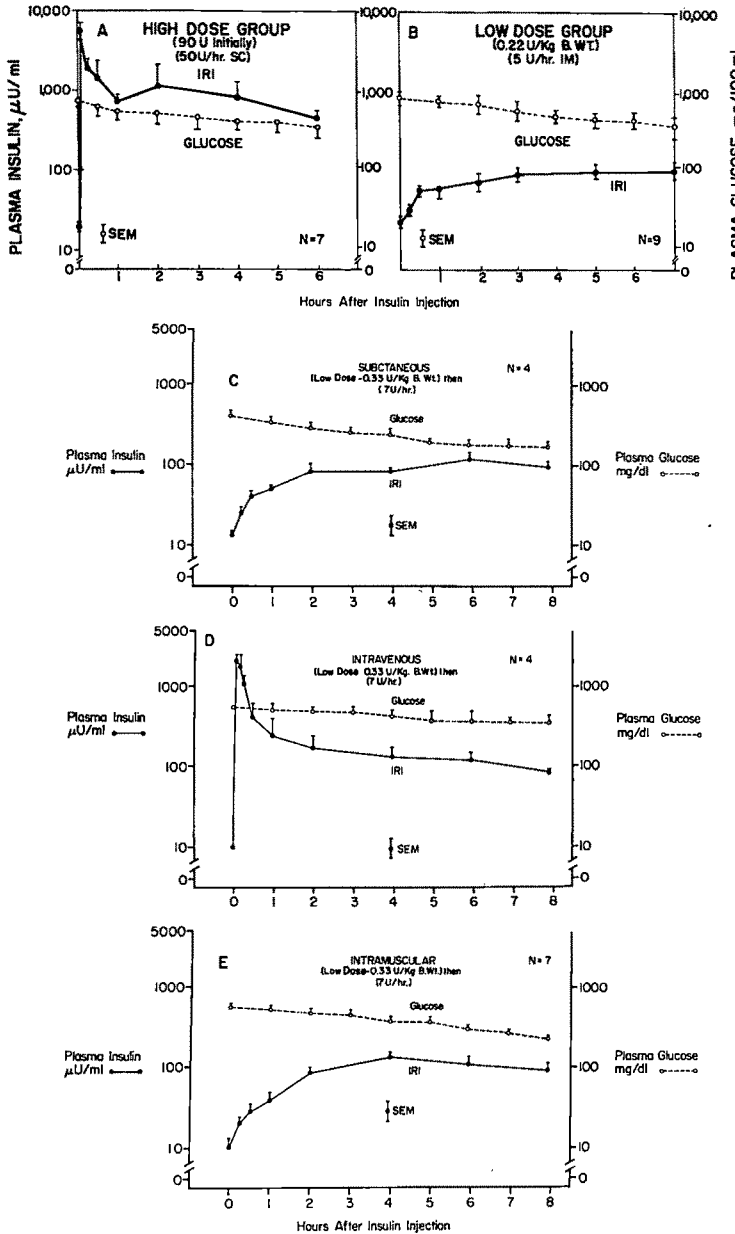


Figure 3 Comparison of high insulin dosage (A) with low insulin dosage group (B) as well as low dose insulin dosage by the SC (C), IV (D), and IM (E) routes in plasma IRI and its glucose lowering effect in DKA patients previously untreated with insulin. Data from Reference 10 (for A and B) and Reference 13 (for C, D, and E).

COMPARATIVE EFFECTIVENESS OF INITIAL PRIMING DOSE OF INSULIN

In the treatment of DKA with low dose insulin, a priming dose of insulin may not be necessary for optimal recovery if continuous IV infusion is used (12). The need for an additional IV bolus to the initial IM and SC loading dose in intermittent IM and SC low dose regimens is not clear. This question was raised when our first studies (Protocols 1 and 2) revealed that in the low dose IM groups 46% and 33% of the patients respectively required repeating priming doses of insulin, whereas only 13% of the high dose IM and SC group and low dose IV groups required the second priming dose (4, 13, 14). To elucidate the mechanism of the delayed initial response, various biochemical and clinical parameters of patients in Protocol 1 were analyzed; these are presented in Table 7. That no significant differences

Table 7 Initial clinical and biochemical profile of 48 patients in Protocol 1 who responded to single loading dose of insulin (responders) vs those who required two or more loading doses of insulin (slow responders)^a

	Responders	Slow responders	Probability
Number of patients	34 ^b	14 ^c	
Age (years)	37.8 ± 3.1	43.6 ± 3.9	>0.05
BP systolic (mm Hg)	126 ± 4	129 ± 8	>0.05
BP diastolic (mm Hg)	74 ± 2	82 ± 5	>0.05
Pulse (beats/min)	109 ± 2	108 ± 5	>0.05
Temperature (°F)	98.1 ± 0.2	97.8 ± 0.3	>0.05
BUN (mg/dl)	31 ± 4	36 ± 5	>0.05
HCO ₃ ⁻ (meq/liter)	8.4 ± 0.5	6.7 ± 0.6	>0.05
pH	7.15 ± 0.02	7.11 ± 0.03	>0.05
K ⁺ (meq/liter)	5.5 ± 0.2	6.1 ± 0.3	>0.05
Na ⁺ (meq/liter)	138 ± 1	135 ± 1	>0.05
Ketone bodies (mM)	13.5 ± 0.4	15.0 ± 1.2	>0.05
Lactate (mg/dl)	20.9 ± 1.8	28.0 ± 3	0.046 ^d
Pyruvate (mg/dl)	0.89 ± 0.13	0.93 ± 0.13	>0.05
IRI (μU/ml) ^e	20.8 ± 1.9	25.0 ± 5.0	>0.05
Glucose (mg/dl)	655 ± 57	682 ± 52	>0.05
Cortisol (μg/dl)	71 ± 7	81 ± 4	>0.05
Glucagon (pg/ml)	423 ± 48	689 ± 131	0.023 ^d

^a A drop from initial plasma glucose of 10% or greater after one hour was considered a response to insulin therapy. Note: Mental status, precipitating factors, obesity, race, and sex were not significantly different by chi square analysis.

^b Includes 13 (54%) of 24 patients in the low dose group and 21 (87%) of 24 patients in the high dose group.

^c Includes 11 (46%) of 24 patients in the low dose group and 3 (13%) of 24 patients in the high dose group.

^d These values became nonsignificant when high and low dose insulin therapy were analyzed separately.

^e Obtained on 16 patients previously untreated with insulin.

(except for lactate and possibly glucagon) were apparent suggests two possibilities: (a) the initial priming IV dose of insulin is necessary to prevent delayed response, or (b) other factors not studied must be contributing to the delayed response. The former hypothesis is strengthened by the fact that in subsequent studies where the only significant modification in our protocol was inclusion of the IV loading dose (Protocol 3), the number of slow responders was reduced to 13% (14), which is identical to the high dose group (Table 7). This suggests that the initial delay in response was a function of the route rather than the dose of insulin.

PEDIATRIC STUDIES

Since the report of Alberti et al in 1973 (8) and our own report in 1976 (10) regarding the efficacy of low dose insulin in treatment of DKA patients, many studies have demonstrated a similar effectiveness of low dose insulin in pediatric patients (17-19). However, randomized studies in this group have been limited. Therefore, Burghen and co-workers investigated the effectiveness of low dose and high dose infusion of insulin in a group of pediatric patients by a randomized protocol (15). These studies demonstrated that an initial low dose infusion may not provide as great a drop in plasma glucose as a high dose, but the time for plasma glucose to reach 250 mg/dl is within six hours in both groups. Of interest was the incidence of hypoglycemia (2 of 16) and hypokalemia (3 of 16) with low dose compared to 12 of 16 and 10 of 16 respectively in the high dose group (15).

EFFECTIVENESS OF LOW DOSE PROTOCOL

Hawkes & Kitabchi (16) used a simplified regimen in a community hospital without the vigorous supervision necessitated by research protocols and showed that low dose insulin therapy of DKA is a practical alternative to conventional therapy.

FLUID AND ELECTROLYTE THERAPY IN DIABETIC KETOACIDOSIS

From balance studies in DKA, deficits of fluid and electrolytes, particularly sodium and potassium, are well described (1, 20) and result principally from renal wastage of these ions by osmotic diuresis and as buffering cations of ketone bodies (20). As insulin directly stimulates renal tubular reabsorption of sodium in man, insulin deficiency may in part be an additional mechanism of the deficits (21).

Prompt and adequate fluid and electrolyte replacement constitutes a critical part of the overall therapy of diabetic ketoacidosis. There is neither

uniformity nor are there established guidelines as to the type of fluid replacement with which to commence or maintain therapy. Plasma sodium concentration in diabetic ketoacidosis is usually reduced or within normal limits (5) in spite of balance studies that indicate hypotonic fluid depletion, i.e. loss of water in excess of solute and loss of cations in excess of chloride (20).

Isotonic or Hypotonic Saline Therapy

The presence of hyponatremia (less than 135 meq/liter) before or especially during treatment may serve to identify individual patients predisposed to the development of cerebral edema, a fact particularly well documented in pediatric cases (22). Thus, Alberti & Hockaday (1) have advocated the sole use of isotonic saline (154 mM sodium chloride) at recommended rates to offset the falling extracellular osmolality as blood glucose is shifted intracellularly following the administration of insulin. This prevents a rapid reversal of osmotic gradients, overexpansion of the intracellular compartment, and the development of cerebral edema. Mild hypernatremia (145–155 meq/liter) may occur, but if the serum concentration exceeds 155 meq/liter, a change at that time to either 77 mM sodium chloride or 5% glucose intravenously, or simply water by mouth, when appropriate, may be made.

One consequence of the use of isotonic saline, first documented by Kydd in 1933 (23), is the development of a hyperchloremic metabolic acidosis, which results from dilution by fluid and displacement by chloride ions of bicarbonate ions in the expanded extracellular fluid. This may occur when the ketosis clearly has been reversed by insulin therapy and is recognized by low bicarbonate concentration and a normal "anion gap" [$\text{Na} - (\text{Cl} + \text{HCO}_3^-) = 8\text{--}12 \text{ meq/liter}$] (24). This acid-base disturbance appears to be transient, reversible, and of no known adverse clinical consequence to the patient (23, 25). Nevertheless, since isotonic saline contains an unphysiological concentration of chloride with respect to the extracellular fluid concentration, hyperchloremia of significant degree ($> 115 \text{ meq/liter}$) can be avoided by using less isotonic solutions (26). In protocols that define control of ketoacidosis as a bicarbonate concentration of at least 15 meq/liter, the persistence of a low bicarbonate, associated with the hyperchloremia, tended to prolong predetermined recovery rates when the hyperketonemia was reversed. It may seem reasonable to initiate therapy with two liters of 154 mM saline given rapidly over the first 1–3 hours, to continue with the same at a slower rate, and to switch thereafter to 77 mM saline if the sodium concentration rises above 145 meq/liter.

An additional obvious complication of vigorous saline therapy includes overhydration leading to cardiac failure. This maybe avoided by the appropriate use of central venous pressure monitoring in any patient with sus-

pected or known cardiovascular or renal disease. Clearly, repeated careful clinical and biochemical assessment of fluid and electrolyte status during treatment of DKA cannot be overemphasized.

Hyperchloremic Acidosis

In some cases, hyperchloremic metabolic acidosis in the first few days following recovery from DKA is due specifically to defects in acidification in the proximal and/or distal renal tubules (27). Further, in poorly controlled diabetes with ketonemia short of frank ketoacidosis, ($\text{pH} > 7.30$; $\text{HCO}_3^- > 15$ meq/liter), hyperchloremic acidosis and markedly increased fractional excretion of ketoacids with potential bicarbonate wastage has been described in association with other renal tubular defects such as phosphaturia and aminoaciduria (28). The exact incidence of such defects in diabetes mellitus is not known, since large series have not been studied, but some authors report that it is high (27). The presence of renal tubular acidosis during recovery from diabetic ketoacidosis may be investigated, not only by more conventional measurements of plasma bicarbonate, blood gas analyses, minimal urine pH, titratable acid, and ammonia excretion, but by sequential measurements of the gradient between the urine PCO_2 and arterial PCO_2 tensions after oral sodium bicarbonate loading. This latter test was advocated by Halperin and co-workers (29) as a useful indicator of distal tubular hydrogen ion secretory capacity.

Effects of Hydration

Rehydration with 77 or 154 mM sodium chloride solution prior to insulin administration is frequently associated with a fall in blood glucose and ketone concentrations greater than can be accounted for by dilution of extracellular fluid alone (1). The factors other than hemodilution that account for this phenomenon have not been fully studied, but in part may result when, because of improved perfusion, residual circulating insulin reaches insulin-sensitive tissues such as skeletal muscle. The initial mean plasma glucose concentration may fall by about 18% after a mean of 1.5 liters of fluid is given over several hours prior to insulin injection. There was no appreciable change in serum acetone, pH, or potassium concentration (30). If insulin is given at the same time as rehydration is commenced, a fall in plasma or blood glucose levels of 10% or more after one hour cannot be attributed solely to an insulin effect, and the glucose must be measured one or two hours thereafter to confirm the sustained drop in glucose.

Potassium Therapy

The timing of potassium supplementation once insulin has been given and the serum potassium starts falling is important to prevent potentially life-

threatening hypokalemia. The initial serum potassium concentration is elevated in most cases (Table 6), and potassium infusion may commence once the potassium level has reached the upper half of the normal range (4.5–5.5 meq/liter). Potassium must be given from the start of insulin therapy with a normal, and particularly a reduced, initial potassium concentration. In the latter case, frequent monitoring of serum potassium concentration and the electrocardiogram are mandatory; there may be some advantage to using low dose insulin therapy in these cases (10), since the potassium supplement may be less and, therefore, easier to administer than in high dose insulin regimens (5). The potassium may be administered either as chloride or neutral phosphate solution or mixtures (13) intravenously, or orally when appropriate. However, if KCl alone is used, the intracellular shift of potassium cation may leave the chloride ion free in the extracellular fluid and contribute to a transient hyperchloremia as noted above.

Bicarbonate Therapy

Alkali therapy continues to be a controversial point as a number of reasons for withholding alkali in DKA therapy have been advanced. Alkali has been used in DKA for decades based on the empiric reasoning that if morbidity and mortality were associated with acidosis, then its correction by adding alkali to the bloodstream would be beneficial. This, however, could be fallacious reasoning as the elevated H^+ ion concentration present in the extracellular fluid is a reflection of the more basic intracellular metabolic alterations, not the cause. The functions of critical intracellular processes for the most part, are well protected from fluctuations in the extracellular environment. Obviously, there is a balance (albeit deteriorating) of compensatory mechanisms in DKA that exert a stabilizing effect on the intracellular milieu. The rapid "correction" of acidosis by administration of alkali may very well upset this balance, to the detriment of the critical intracellular processes, and the patient as a whole.

The alkali-in-DKA controversy flourishes because studies on this subject have been conflicting and inconclusive as noted in the well-documented summaries of Kaye (31) and Matz (32). Even the putative harmful effects of acidosis per se on the cardiovascular system and myocardial function is an area fraught with insufficient clinical studies arriving at conflicting conclusions as exemplified by papers of Bierholm et al (33), Smith et al (34), and Gonzalez et al (35).

The following observations offer the most convincing arguments against vigorous correction of the metabolic acidosis in DKA with alkali therapy:

1. Because of differential permeability of CO_2 and HCO_3^- across the blood brain barrier, the rising arterial PCO_2 that accompanies "correction"

of the acidosis by alkali administration readily gains access to the CSF, while the HCO_3^- lags. This results in the "paradoxical CNS acidosis" first described by Posner & Plum (36). Ohman et al (37) and Assal et al (38) substantiated this finding but found that the acidosis was transient and did not compromise CNS function. In like manner, the differential permeability across cell membranes has been theorized to lead to "paradoxical" intracellular acidosis as well (39).

2. A more thoroughly investigated area that offers an indictment against alkali therapy in mild acidosis in DKA has been championed by Standl & Ditzel (40), Alberti et al (41), and other investigators; it relates to an exacerbation of the already compromised oxygen-releasing capacity of blood at the tissue level due to low 2,3-diphosphoglycerate (2,3-DPG) levels associated with treatment phase of DKA. The counter-balancing effect of oxygen hemoglobin dissociation conferred by the Bohr effect in acidosis is removed when the acidosis is vigorously corrected by alkali. This shifts the O_2 dissociation curve to the left and increases affinity of hemoglobin for oxygen with aggravation of tissue hypoxia.

Phosphate Therapy

Phosphate therapy, although not universally employed in DKA, is not a controversial subject—just poorly understood by many practitioners. It has been well documented that serum phosphate levels fall from initial normal or elevated levels during therapy of DKA (42, 43). The depressed phosphate levels are of biochemical and physiological consequence, because they are intimately related to oxygen delivery to peripheral tissues and to production of high energy phosphate compounds. Insulin therapy itself may be responsible for the drop in serum phosphorous observed in the treatment of DKA; Ditzel et al (44) relate the drop to uptake of phosphorous into the cell and use in insulin-related phosphorylations, as well as to the production of high energy phosphate compounds. There is also an association of hypophosphatemia and hypokalemia, which may arise from an acquired renal tubular absorptive defect in DKA (28, 45) as already mentioned.

SUMMARY AND PROSPECTS

Although the overwhelming evidence from numerous medical centers around the world suggests that low dose insulin regimen is effective in the management of DKA, well-controlled randomized studies to compare efficacy of low dose to that of the conventional methods of insulin therapy are relatively few. In spite of the fact that these limited studies (total of six) are somewhat well controlled, they contain shortcomings, including inconsis-

tent criteria for defining DKA, inadequate initial biochemical values, and nonuniform criteria of acceptable response, all of which make comparison between studies difficult.

Of even greater irony is the persistent utterance without a sound scientific basis of certain groups of diabetologists on the superiority of the conventional method of therapy (46). In fact, as limited as comparative studies are, none claim superiority of the conventional method over the low dose regimen (Table 1). The latter method appears to be freer of the potential complications that usually attend the high dose protocol, such as hypoglycemia and hypokalemia. The low dose regimen, regardless of the route of therapy, appears to be more reliable, practical, and simpler than a high dose regimen. These statements are based not only on the reports of studies from other medical centers, but also on the experimental results obtained on the treatment of more than 150 DKA patients in this medical center. The environment for such studies ranged from a most regimented and controlled setting to that of the more practical and realistic situations that prevail in nonspecialized community hospitals here and elsewhere (47).

In spite of these successful reports, there remain major gaps in our knowledge concerning the management of DKA patients. These may be posed as the following questions:

1. What is the fate of injected insulin in DKA patients, and is there a difference in the metabolism of high vs low dose insulin if similar target organ response is achieved?
2. What is the extent, if any, of insulin resistance in DKA, and what parameter(s) is responsible for this phenomenon?
3. Is the use of phosphate therapy essential, and if so, will the immediate salutary effect be through improvement of the level of glycolytic intermediate 2,3-DPG, high energy phosphate compounds, or other factors modulating the insulin effect?
4. What are the merits or demerits of alkali therapy in DKA when only moderate acidosis (i.e. pH between 7.0 and 7.3) exists?
5. Should DKA patients be treated with isotonic saline alone? What are the advantages and/or disadvantages of such a therapy? And finally, a more fundamental question:
6. What is an appropriate rate of recovery of hyperglycemia, ketonemia, or acidosis? Is there in fact a demonstrated deleterious effect in DKA patients exposed to ketonemia for longer periods than say 6 hours? 8 hours? 12 hours? or 24 hours?

Although opinions vary with each diabetologist regarding the above questions, it is safe to assume that the majority of these opinions are based on limited clinical experience and not on solid well-controlled experimental

data (48). A plea is therefore made for continued diligent investigation, not only on documentation of efficacy of one method of insulin therapy over the other, which by now is a minor point (see below), but on more fundamental issues alluded to in the previous paragraph.

It is axiomatic to say that any amount of insulin by any route is effective in the therapy of DKA as long as (a) circulating IRI concentration ranges from 20–200 $\mu\text{U/ml}$; (b) adequate replacement of fluid and electrolytes is provided; and (c) meticulous attention is paid to the clinical conditions of the patient by the attending physician (1–5, 47–50).

The fundamental question and the important challenge is therefore to provide the scientific community with a better understanding on the pathogenesis of DKA, and a scientific basis for its management through proven, well-controlled investigations performed in an environment free from preconceived notions and “authoritative” pronouncements (48, 51, 52).

ACKNOWLEDGMENTS

The authors gratefully acknowledge the excellent assistance and cooperation of the nursing staff and laboratory personnel of the Clinical Research Center, as well as the assistance of the house staff of the University of Tennessee and City of Memphis Hospitals in the studies of the authors. The secretarial assistance of Ms. Karen Willard, Mable Mackey, and Rebecca Hall are also gratefully acknowledged.

Literature Cited

1. Alberti, K. G. M. M., Hockaday, T. D. R. 1977. Diabetic coma: reappraisal after five years. *Clin. Endocrinol. Metab.* 6:421–55
2. Alberti, K. G. M. M. 1977. Low dose insulin in the treatment of diabetic ketoacidosis. *Arch. Intern. Med.* 137: 1367–76
3. Padilla, A. J., Loeb, J. N. 1977. Low dose versus high dose insulin regimens in the management of uncontrolled diabetes, a survey. *Am. J. Med.* 63:843–48
4. Kitabchi, A. E. 1978. Treatment of diabetic ketoacidosis with low dose insulin. *Adv. Intern. Med.* 23:115–35
5. Kriesberg, R. A. 1978. Diabetic ketoacidosis: new concepts and trend in pathogenesis and treatment. *Ann. Intern. Med.* 84:681–95
6. Smith, K., Martin, H. E. 1954. Response of diabetic coma to various insulin dosages. *Diabetes* 3:287–95
7. Shaw, C. E., Hurwitz, G. E., Schumkier, M., Brager, S. H., Bessman, S. P. 1962. A clinical and laboratory study of insulin dosage in diabetic acidosis: comparison with small and large doses. *Diabetes* 2:23–30
8. Alberti, K. G. M. M., Hockaday, T. D. R., Turner, R. C. 1973. Small doses of intramuscular insulin in the treatment of diabetic “coma.” *Lancet* 2:515–22
9. Alberti, K. G. M. M., Hockaday, T. D. R. 1972. Diabetic coma: serum growth hormone before and during treatment. *Diabetologia* 9:12–19
10. Kitabchi, A. E., Ayyagari, V., Guerra, S. M. O., Medical House Staff. 1976. The efficacy of low dose versus conventional therapy of insulin for treatment of diabetic ketoacidosis. *Ann. Intern. Med.* 84:633–38
11. Heber, D., Molitch, M. E., Sperling, M. A. 1977. Low dose continuous insulin therapy for diabetic ketoacidosis. *Arch. Intern. Med.* 137:1377–80
12. Piters, K. M., Kumar, D., Pei, E., Bessman, A. N. 1977. Comparison of continuous and intermittent intravenous insulin therapies for diabetic ketoacidosis. *Diabetologia* 13:318–21

13. Fisher, J. N., Shahshahani, M. N., Kitabchi, A. E. 1977. Diabetic ketoacidosis: low dose insulin therapy by various routes. *N. Engl. J. Med.* 297: 238-41
14. Sacks, H. S., Shahshahani, M. N., Kitabchi, A. E., Fisher, J. N., Young, R. T. 1978. Similar responsiveness of diabetic ketoacidosis to low dose insulin by intramuscular injection and albumin-free infusion. *Ann. Intern. Med.* In press
15. Burghen, G. A., Kitabchi, A. E., Fisher, J. N., Etteldorf, J. E. 1978. Low dose versus high dose infusion of insulin in the treatment of diabetic ketoacidosis in children. Submitted
16. Hawkes, J., Kitabchi, A. E. 1978. Efficacy of low dose insulin in treatment of diabetic ketoacidosis in a non specialized community hospital. Submitted
17. Martin, M. M., Martin, A. L. 1976. A continuous low dose infusion of insulin in the treatment of diabetic ketoacidosis in children. *J. Pediatr.* 89:560-66
18. Drop, S. L. S., Duval-Arnold, B. J. M., Gover, A. E., Hersh, J. H., McEvey, P. T., Knowles, H. C. 1977. Low dose intravenous insulin infusion versus subcutaneous insulin injection: a controlled comparative study of diabetic ketoacidosis. *Pediatrics* 59:733-38
19. Edwards, G. A., Kohaut, E. C., Wehring, B., Hall, L. L. 1977. Effectiveness of low-dose continuous insulin infusion in diabetic ketoacidosis. *J. Pediatr.* 11: 701-5
20. Arieff, A. I. 1976. Kidney, water and electrolyte metabolism in diabetes mellitus. In *The Kidney*, ed. B. M. Brenner, F. C. Rector, pp. 1257-96. Philadelphia: Saunders
21. DeFronzo, R. A., Cooke, C. R., Andres, R., Faloona, G. R., Davis, P. J. 1975. The effect of insulin on renal handling of sodium, potassium, calcium and phosphate in man. *J. Clin. Invest.* 55:845-55
22. Duck, S. C., Weldon, V. V., Pagliara, A. S., Haymond, M. W. 1976. Cerebral edema complicating therapy for diabetic ketoacidosis. *Diabetes* 25:111-15
23. Kydd, D. 1933. Salt and water in the treatment of diabetic acidosis. *J. Clin. Invest.* 12-1169-83
24. Smithline, N., Gardner, K. D. 1976. Gaps—anionic and osmolal. *J. Am. Med. Assoc.* 1594-97
25. Danowski, T. S., Winker, A. W., Peters, J. P. 1946. Salt depletion, peripheral vascular collapse and treatment of diabetic acidosis. *Yale J. Biol. Med.* 28: 405-17
26. Sprague, R. G., Power, M. H. 1953. Electrolyte metabolism in diabetic ketoacidosis. *J. Am. Med. Assoc.* 151: 970-76
27. Giammarco, R., Goldstein, M. B., Halperin, M. L., Stinebaugh, B. J. 1975. Renal tubular acidosis during therapy for diabetic ketoacidosis. *Can. Med. Assoc. J.* 112:463-66
28. Hammeke, M., Bear, R., Lee, R., Goldstein, M., Halperin, M. 1978. Hyperchloremic metabolic acidosis in diabetes mellitus. A case report and discussion of the pathophysiologic mechanisms. *Diabetes* 27:16-20
29. Halperin, M. L., Goldstein, M. B., Haig, A., Johnson, M. D., Stinebaugh, B. J. 1974. Studies on the pathogenesis of type I (distal) renal tubular acidosis as revealed by the urinary PCO_2 tensions. *J. Clin. Invest.* 53:669-77
30. Sacks, H. S., Shahshahani, M. N., Kitabchi, A. E., Gray, T. L. 1977. Diabetic ketoacidosis: effect of intravenous fluid prior to insulin therapy. *Clin. Res.* 25:4, 624A
31. Kaye, R. 1975. DKA—the bicarbonate controversy (editorial). *J. Pediatr.* 87: 156-59
32. Matz, R. 1976. Diabetic acidosis: rational for not using bicarbonate. *NY State J. Med.* 76:1299-1302
33. Beierholm, E., Grantham, N., O'Keete, D., Laver, M., Dagget, W. 1975. Effects of acid-base changes, hypoxia, and catecholamines on ventricular performance. *Physiology* 228:1555-61
34. Smith, N., Corbascio, A. 1966. Myocardial resistance to metabolic acidosis. *Arch. Surg.* 92:892-97
35. Gonzalez, N., Clancy, R. 1975. Isotropic and intracellular acid-base changes during metabolic acidosis. *Physiology* 228:1060-64
36. Posner, J., Plum, F. 1967. Spinal fluid pH and neurologic symptoms in systemic acidosis. *N. Engl. J. Med.* 277: 605-13
37. Ohman, J., Marliss, E., Aoki, T., Muni-choodappa, C., Khanna, V., Kozak, G. 1971. Cerebrospinal fluid in diabetic ketoacidosis. *N. Engl. J. Med.* 284: 283-89
38. Assal, J., Aoki, T., Manzano, F., Kozak, G. 1974. Metabolic effects of sodium bicarbonate in management of diabetic ketoacidosis. *Diabetes* 23: 405-11
39. Robin, E., Wilson, R., Bromberg, P. 1961. Intracellular acid-base relations

- and intracellular buffers. *NY Acad. Sci. Ann.* 92:539-46
40. Standl, E., Ditzel, J. 1976. The effect of red cell 2,3-diphosphoglycerate changes induced by diabetic ketoacidosis on parameters of the oxygen dissociation curve in man. *Adv. Exp. Med. Biol.* 75:89-95
 41. Alberti, K. G. M. M., Darley, J., Emerson, P., Hockaday, T. D. R. 1972. 2,3-diphosphoglycerate and tissue oxygenation in uncontrolled diabetes mellitus. *Lancet* 2:391-95
 42. Guest, G., Rapoport, S. 1947. Electrolytes of blood plasma and cells in diabetic acidosis and during recovery. *Proc. Diabetes Assoc.* 7:96-115
 43. Nabarro, J., Spencer, A., Stowers, J. 1952. Metabolic studies in severe diabetic ketosis. *Q. J. Med.* 21:225-43
 44. Ditzel, J. 1973. Importance of plasma inorganic phosphate on tissue oxygenation during recovery from diabetic ketoacidosis. *Horm. Metab. Res.* 5: 471-72
 45. Relman, A., Schwartz, W. 1956. The nephropathy of potassium depletion. *N. Engl. J. Med.* 255:195-203
 46. Root, H. F., 1945. The use of insulin and abuse of glucose in the treatment of diabetic coma. *J. Am. Med. Assoc.* 127:557-64
 47. Bendezu, R., Wieland, R. G., Furst, B. H., Mandel, M., Genuth, S. M., Schumacher, P. 1978. Experience with low-dose insulin infusion in diabetic ketoacidosis and diabetic hyperosmolarity. *Arch. Intern. Med.* 138:60-62
 48. Ingelfinger, F. J. 1977. Debates on diabetes. *N. Engl. J. Med.* 296:1228-30
 49. Felig, P. 1974. Insulin: Rates and routes of delivery. *N. Engl. J. Med.* 291: 1021-22
 50. Editorial. 1977. Insulin regimens for diabetic ketoacidosis. *Br. Med. J.* 1:405-6
 51. Madison, L. L. 1976. Low dose insulin: a plea for caution. *N. Engl. J. Med.* 294:393
 52. Kitabchi, A. E. 1977. Insulin therapy in diabetic ketoacidosis. Comments. *Ann. Intern. Med.* 86:110-11

DISSEMINATED INTRAVASCULAR COAGULATION: A REAPPRAISAL¹

◆7329

Robert W. Colman, M.D.

Hematology-Oncology Section, Special Center for Thrombosis Research,
and Temple University Medical Center, Philadelphia, Pennsylvania 19140

Stanley J. Robboy, M.D.

Department of Pathology, Harvard Medical School, and Massachusetts General
Hospital, Boston, Massachusetts 02114

John D. Minna, M.D.

NCI-VA Medical Oncology Branch, Division of Cancer Treatment, National
Cancer Institute, and Washington Veterans Administration Medical Center,
Washington, DC 20422

INTRODUCTION

The purpose of this review is to reappraise the syndrome of disseminated intravascular coagulation (DIC: defibrination syndrome, intravascular coagulation with fibrinolysis syndrome, consumption coagulopathy) with respect to recent advances in the molecular biology of the coagulation system, laboratory diagnosis, initiating or predisposing conditions, pathophysiology, clinical features, organ dysfunction, pathology, relation to liver disease, and therapy. There have been many reviews of this subject and the reader is particularly directed to References 1-17. In addition, recent editions of the major textbooks of medicine and hematology contain chapters on DIC.

This review is based in part on a National Library of Medicine Medlars II search. This search revealed 681 articles on DIC on file, 382 of which coded DIC as the main topic. Between January 1977 and June 1978, 265

¹The US Government has the right to retain a nonexclusive, royalty-free license in and to any copyright covering this paper.

articles on DIC were recorded; of these 129 were in English, and 44 were related to DIC therapy, drug therapy, or complications. We selected the references to review from the overall search, with emphasis on articles published in the past five years.

PATHOGENESIS

DIC is a pathological syndrome resulting from the formation of thrombin, subsequent activation and consumption of certain coagulant proteins, and production of fibrin thrombi. These basic processes reflect the multiple actions of thrombin in the hemostatic system. First, thrombin proteolytically cleaves fibrinogen to produce fibrin monomer, which may complex with fibrinogen or other protein to circulate as soluble fibrin monomer or may polymerize to form fibrin thrombi. Two small fibrinopeptides, A and B, which may cause systemic (18) or pulmonary (19) vasoconstriction, are liberated in the process. Thrombin converts factor XIII (fibrin-stabilizing factor) to an active transamidase, which in turn crosslinks fibrin with covalent amide bonds that render the fibrin more resistant to fibrinolysis (20). Thrombin binds to platelets at low concentrations and initiates shape change, aggregation, and secretion. Aggregated platelets are then removed from the circulation. Those remaining in the circulation and that lack granules are functionally defective both *in vivo* and *in vitro* (21). Thrombin also makes a membrane phospholipoprotein (platelet factor 3) available on the surface of the aggregated platelet. This factor is necessary for the interactions of factors IXa-VIII and Xa-V-prothrombin (22). Thrombin potentiates the coagulation cascade by increasing the activity of factor V, a protein itself necessary for the optimal formation of thrombin. Factor V proteolyzed by thrombin has unstable derivatives that decay rapidly (23), which further decreases factor V levels. A similar transformation occurs when thrombin interacts with factor VIII. Thus, thrombin itself can account for decreased fibrinogen, platelets, and factors II, V, VIII, and XIII in acute DIC. The decrease in other coagulant proteins such as IX and X is due to rapid clearance of activated clotting factors *in vivo* (24). The liver removes activated clotting factors, and the reticuloendothelial system (RES) removes fibrin. However, thrombin infusion appears to temporarily depress RES function (25, 26).

In addition to thrombin, the fibrinolytic system influences manifestations of DIC. Activation of factor XII leads to kallikrein production, which in turn results in the conversion of plasminogen to plasmin (27), a potent proteolytic enzyme capable of digesting fibrin. Endothelial cells contain plasminogen activator, although the nature of the stimulus for its release is unknown. Injured tissues release potent plasminogen activators simulta-

neously with thromboplastin. Thus fibrinolysis almost invariably accompanies thrombin evolution. The balance between these two proteases undoubtedly determines the degree of hemorrhage and/or thrombosis seen clinically. Plasmin destroys factors V, VIII, and XIII, the very factors that thrombin potentiates. Plasmin digestion of fibrinogen results in fibrin degradation products (FDP) that are critical to the pathogenesis and diagnosis of DIC. At short incubation times or low plasmin concentrations, fibrinogen (MW 330,000) is degraded to fragment X (28, 29), which clots slowly. Additional digestion cleaves X to Y (MW 150,000), a potent inhibitor of fibrin polymerization. Finally Y is cleaved asymmetrically to produce a second molecule of D and fragment E (MW 50,000). Thus, the large fragments Y and D are potent inhibitors of fibrin formation, while small fragments may inhibit platelet aggregation. The importance of fibrinolysis as a compensatory mechanism to lyse fibrin clots of DIC is shown by cases in which the fibrinolytic inhibitor epsilon aminocaproic acid (EACA), administered to treat the fibrinolysis, resulted in severe thrombotic complications (30).

LABORATORY DIAGNOSIS

Screening Tests

DIC can be detected by the combination of abnormal prothrombin time (PT), thrombocytopenia, and hypofibrinogenemia. Although the PT may occasionally be prolonged by conditions such as vitamin K deficiency, liver disease, or dilution, it is abnormal (3 sec over control) in over 90% of patients with DIC (10). Of patients with DIC, 70% have fibrinogen levels less than 150 mg/100 ml (10). The normal fibrinogen concentration in the remaining 30% can be explained by noting that both major causes of DIC, sepsis and neoplasia, lead to elevated fibrinogen levels. Thus a fibrinogen level in the normal ranges in these conditions is inappropriately low.

The platelet count, like the PT, is abnormal in over 90% of cases of DIC. However, it must be carefully evaluated because of the wide variety of other causes of thrombocytopenia, including bone marrow depression due to the diseases underlying DIC. We have found another common test, the partial thromboplastin time (PTT), is not very useful in diagnosis. In some normal patients, the PTT may be shorter than normal (31). The wide range of normal, e.g. 22–35 sec, precludes detection of significant changes because of the 13-sec interval. In subacute defibrination (31), only 20% of the PTTs were abnormal.

Abnormalities in all three screening tests (PT, fibrinogen concentration, platelet count) establish the diagnosis of DIC in the absence of dilution or severe hepatic dysfunction. When only two screening tests are abnormal at any given time, confirmatory tests may be required to diagnose DIC.

Confirmatory Tests

The most useful confirmation of DIC is the measurement of FDP. The Thrombo-Wellco test (32) uses latex particles coated with antibodies to FDP fragments D and E. The staphylococcal-clumping test measures fragments X and Y (33), while the tanned-red-cell hemagglutination inhibition test (34) is sensitive to all forms of FDP but can give false negative values in the presence of an anti-D (Rh) antibody (35). All three tests correlate closely. The measurement of FDP by any one of these three methods is useful in the diagnosis of DIC; it is elevated ($>5 \mu\text{g/ml}$) in about 95% of cases. However, mild elevations up to $40 \mu\text{g/ml}$ occur in a wide variety of conditions (33), including hepatocellular disorders, and in the postoperative state (36). Values greater than $40 \mu\text{g/ml}$ almost always indicate DIC.

Other tests are less satisfactory for confirming the diagnosis of DIC. The thrombin time, abnormal in about two thirds of DIC patients, in part reflects the functional effects of FDP, namely the inhibition of thrombin proteolysis or fibrin-monomer polymerization. However, it may be abnormal solely on the basis of hypofibrinogenemia. The euglobulin lysis time, which reflects plasminogen activators after removal of inhibitors, is positive only in a minority of cases (10).

Numerous tests have been devised to measure fibrin-monomer, fibrinopeptides A and B, and end products of fibrinogen polymerization. The better methods are still too complex for routine use and the simple methods are too frequently misleading. Paracoagulation tests with ethanol or protamine (37) purport to measure fibrin monomer and are simple to perform but relatively insensitive and difficult to accurately quantify. Tests used in research studies are potentially adaptable to the clinical laboratory. Fibrin monomer has two moles each of terminal glycine, which is exposed after cleaving of arginine-glycine bonds in the $\text{A}\alpha$ and $\text{B}\beta$ chains of fibrinogen. Methods to quantify these new N-terminals have been developed (38), but have yet to become practical. Agarose gel filtration (39) identifies high molecular weight fibrinogen derivatives representing soluble complexes of fibrin monomers with fibrinogen or fragment X. Unfortunately this test may be too sensitive since it is also positive in a wide variety of conditions (without other signs of DIC) such as ingestion of oral contraceptive drugs, venous thrombosis, nephrotic syndrome, polycythemia vera, coronary artery disease, and hyperbetalipoproteinemia. Similar considerations apply to radioimmunoassay of fibrinopeptide A (40) since elevations have also been reported in pulmonary embolism, thrombophlebitis, septicemia, metastatic carcinoma, systemic lupus erythematosus, renal-transplant rejection, and aortic aneurysm (41). Once formed, thrombin rapidly combines with antithrombin to form an inactive complex. Antithrombin III decreases in all causes of DIC studied, but it is also decreased in hepatic failure (42).

Methods to specifically detect a thrombin-antithrombin III complex as opposed to antithrombin III itself are being developed. Fibrinogen half-life is markedly shorter than normal in patients with DIC. However, shortened fibrinogen survival occurs in many conditions other than DIC and the length of the determination (several days) limits its practical application.

PATHOPHYSIOLOGY

Intrinsic Coagulation System and the Kallikrein-Kinin System

Activation of Hageman factor, which leads to coagulation, kinin formation, and fibrinolysis, occurs in many of the cases of DIC secondary to infection, especially in gram-negative septicemia. Kallikrein formation releases bradykinin and manifests itself clinically as hypotension. During hypotensive septicemia levels of Hageman factor, prekallikrein and kallikrein inhibitors decrease (43), and bradykinin level rises (44). In contrast, the DIC associated with malignancy was not associated with Hageman factor activation nor were changes in prekallikrein or kallikrein inhibitors found (45).

Complement

Complement 3 (C3) is depressed in DIC owing to a wide variety of causes (46). This decrease involves at least the classical pathway, since C4 is also decreased in parallel to C3 (47). In four patients with DIC caused by Rocky Mountain spotted fever, C3 was decreased although an exact temporal correlation to platelets or fibrinogen was not shown (48). In malaria and DIC the decrease of C3 correlated with the severity of the thrombocytopenia, while a parallel consumption of C4 suggested that antibody-antigen activation occurred (49). In rabbits DIC induced by thromboplastin or thrombin is associated with a decrease in C3 (50). However, if C3 is depleted by cobra venom, endotoxin will still induce DIC in rabbits with concomitant hypotension and leukopenia (51). Moreover, endotoxin-induced DIC was not altered in rabbits that were congenitally deficient in C6 and had defective platelet-factor-3 availability (52). Thus, it appears that consumption of C3 and C4 is associated with DIC but not necessary for its induction.

Hematologic Interactions in DIC

In the rabbit, the leukocyte, not the platelet, appears to be essential for the production of DIC (53). Animals in which platelets were depleted with a specific antiserum (54) or neuraminidase (55) developed DIC after endotoxin challenge, while neutropenia induced by nitrogen mustard totally inhibited DIC (55, 56). Niemetz (57) has shown that rabbit leukocytes have procoagulant activity resembling thromboplastin.

The role of the red cell in the pathogenesis of DIC has also been investigated. Spector & Crosby (58) produced experimental hemoglobinuria in humans with autologous freeze-thawed blood, but failed to induce sub-clinical DIC. It seems likely that the clinical DIC observed with severe transfusion reactions is probably secondary to the consequences of antibody-antigen complex formation rather than the effect of hemolysis.

While thrombocytopenia is a cardinal feature of DIC, the process also produces a qualitative defect in platelets that resembles an acquired storage pool defect (59). It is characterized by decreased aggregation with ADP and diminished release of ADP and platelet factor 4. Further studies revealed (60) defective second phase aggregation with epinephrine and collagen, decreased ADP and ATP content with elevated ATP:ADP ratio, and low content and uptake of serotonin, all of which are characteristic of the "storage pool" defect. Since Reimers et al (61) showed that platelets degranulated with thrombin have a normal lifespan, it seems likely that the qualitative platelet abnormalities in DIC are related to exposure of circulating platelets to thrombin with subsequent degranulation.

CONDITIONS INITIATING OR PREDISPOSING PATIENTS TO DIC

DIC is a sign of a disease process that may have many different causes. The most common cause of DIC is infection, a subject that has recently been reviewed (13, 62). Bacterial infections, especially those associated with gram-negative endotoxin-containing organisms, are probably most common. However, DIC also has been reported in typhoid fever (63), where endotoxin cannot be demonstrated by limulus assay in the blood. Gram-positive sepsis due to staphylococci, streptococci, or pneumococci, *Rickettsia* (including Rocky Mountain spotted fever), and fungi (including *Histoplasma* and *Aspergillus*) have been responsible for DIC. Viruses also trigger DIC, the most consistent being the epidemic arboviruses such as those that produce dengue hemorrhagic fever. A recent study of 29 patients with dengue showed a correlation among clinical severity including shock, laboratory evidence for DIC, and shortened half-life of fibrinogen (64).

Localized endothelial injury if severe may also manifest itself as DIC. Giant hemangiomas (65, 66), aortic aneurysms (67), and angiography (68), which presumably causes mechanical injury to endothelium, are reportedly associated with subacute consumption coagulopathy, as is documented by fibrinogen survival studies. Catheterization of vessels during DIC frequently causes thrombosis in the traumatized vessel.

The second most common cause of DIC is neoplasia, particularly leukemia and pancreatic and prostatic cancers, though a wide variety of tu-

mors appear to predispose to the development of DIC. The triggering mechanism is thought to be release of coagulant proteins from the tumor (69). Tissue thromboplastic activity is higher in malignant tissue than in the normal counterparts, but more specific coagulants, e.g. factor-X-activating protein, may also be involved. One large study of premalignant and malignant disease reviewed the histories of 89 patients (6), and another study reviewed 10 personal cases augmented with 172 from the literature (5). In the former, the retrospective diagnosis reemphasized the occurrence of bleeding in about three quarters of the individuals, while arterial and venous thrombosis and pulmonary embolism comprised the other fourth. The latter study reviewed Trousseau's syndrome. As expected, thrombotic complications such as migratory thrombophlebitis, arterial embolism, and nonbacterial thrombotic endocarditis were more frequent than hemorrhage. Owen & Bowie (11) point out that chronic DIC is usually clinically mild and difficult to diagnose. The typical features of DIC may be less striking in compensated intravascular coagulation with mildly depressed platelets and elevated FDP, but with normal fibrinogen and coagulation factors. Concentrations of thromboplastin or other procoagulant enzymes have not been demonstrated in the blood, but temporary exacerbation of DIC with cytotoxic therapy in leukemia, and hormonal therapy or instrumentation (70) in prostatic carcinoma, provide circumstantial evidence that released tumor products are important. Nearly all tumors reported to be associated with DIC were metastatic. In contrast to the thrombotic events observed in many carcinomas, acute leukemia when symptomatic usually presents as a hemorrhagic disorder. Gralnick et al (71) also distinguish a subclinical compensated state with abnormal laboratory findings but not clinical hemorrhage from an even milder variant with shortened fibrinogen survival as the sole abnormality. The latter variant may occur in most leukemia patients. Promyelocytic leukemia is almost always associated with acute heparin-resistant DIC. Giant lysosomal granules may contain the coagulant material (72). Myeloproliferative disorders such as chronic myelogenous leukemia (73) and polycythemia vera (74) are rare causes of DIC. Another example in which the coagulant substances enter the circulation is snake-bite. For example, the bite of *Echis carraitus* results in classic DIC because the venom directly converts prothrombin to thrombin (75). Some snake bites, for example those of the timber rattlesnake, *Crotalus horridus horridus*, result in pseudo DIC (76) with hypofibrinogenemia due to a defibrinating enzyme (77) and decreased platelets due to a platelet-aggregating protein (78). The coagulation factors remain at normal levels since no thrombin is formed.

Several classic examples of DIC precipitated by thromboplastic or procoagulant substances that enter the circulation are related to pregnancy

or head injury (79-81). DIC in pregnancy may arise locally, as in abruptio placentae or hypertonic saline abortion, or systemically, as in eclampsia. DIC may be an acute event, e.g. amniotic fluid embolism, or a chronic process as in retention of a dead fetus. An excellent clinical review (82) and a good discussion of laboratory changes (83) are available. Abruptio placentae occurs in about 2% of pregnancies and it is the most common obstetric cause of DIC. It accounts for half the maternal deaths caused by hemorrhage. In 25% of cases where the separation of the placenta from the endometrium involves more than 50% of the surface area, coagulation changes can be detected. DIC also develops in as many as 25% of patients in whom a dead fetus has been retained for five weeks. Laboratory abnormalities are often present prior to the bleeding, and hemorrhage may first occur during the dilatation and curettage that cures the condition. Amniotic fluid embolism, while rare, presents as a medical emergency because of the sudden development of shock and cyanosis and severe hemorrhage. The DIC is triggered by embolization of thromboplastic material in the amniotic fluid into the maternal venous circulation, while the clinical manifestations are accentuated by hypotension and acute respiratory distress syndrome. Eighty per cent of patients die. Prior to legalization of abortion, sepsis was common and frequently resulted in DIC. Abortion induced by hypertonic saline or hyperosmotic glycine and urea (84) results in the laboratory finding of DIC, while abortion induced by prostaglandin E_2 or F_{2a} does not.

Necrotizing enterocolitis is another syndrome that may trigger DIC through release of tissue procoagulants from the damaged bowel. The condition occurs in 5% of premature neonates and is characterized by patchy intestinal necrosis with or without frank sepsis. Ileus, vomiting, and intestinal bleeding may lead to perforation, peritonitis, and death. Over 85% of patients had low platelet counts (85) and half of these patients had DIC. The occurrence of DIC is a bad prognostic sign, being four times as common in nonsurvivors as in survivors (86). Pulmonary parenchymal damage may be the underlying mechanism in the DIC that occurs during freshwater drowning (87).

DIC can be triggered by tissue thromboplastin release resulting from surgery. For example, during transurethral prostatectomy, irrigation transports tissue fluids into the circulation (and if distilled water is used, causes hemolysis of viable cells). Prostatic tissue is rich in proteolytic enzymes and plasminogen activators and therefore operations are commonly associated with laboratory evidence for DIC and occasionally with clinical manifestations.

Acute DIC is a recognized complication of treatment of liver disease with commercial concentrates of factors II, VII, IX, and X (88-90) and is due

to the inadequacy of the damaged liver to synthesize antithrombin III and to clear activated clotting factors such as factor Xa and thrombin. In identifying the various processes that are temporally related to the episode of DIC, it is helpful to identify not only the underlying disease but also the precipitating factors.

Laboratory evidence of DIC typically follows cardiac arrest, but overt bleeding and thrombosis are rare (91). The rare (92) occurrence of DIC following extracorporeal circulation is probably due to the fact that patients are receiving heparin at high doses despite the predisposing condition of shock, acidosis, and hepatic failure (93). In contrast, DIC is common in patients who develop acute renal failure associated with cardiogenic shock (94).

CLINICAL CONDITIONS

Acute DIC

There are two distinct modes of presentation of DIC: acute and chronic. The laboratory and clinical features of acute DIC develop rapidly over a period of a few hours to days. The patient is usually critically ill from the precipitating event as well as from the DIC. The patient often presents with a multisite bleeding diathesis that can range from oozing to catastrophic life-threatening hemorrhage. Dermatologic manifestations of DIC are common (95). Thrombotic complications are unusual in acute DIC; 8% in one series (10) of patients with acute DIC had thrombotic complications.

The mortality associated with acute DIC is high, ranging from 54–67% (7, 10). Siegal et al (7) found the mortality directly related to age, the number of clinical manifestations of DIC, and the severity of the laboratory manifestations.

Chronic DIC

The clinical and laboratory features of chronic DIC wax and wane over periods of months and the patient is usually not critically ill. The underlying diseases associated with chronic DIC are malignancy, connective tissue disease, renal disease, and congenital hemangiomas. While episodes of bleeding can occur, thrombotic complications predominate. The largest numbers of patients with chronic DIC are those with an underlying malignancy. Sack et al (5) elegantly reviewed the literature and analyzed 182 cases of chronic DIC and malignancy. They found migratory thrombophlebitis in 53%, a single episode of thrombophlebitis in 62%, bleeding in 41%, arterial emboli in 25%, nonbacterial thrombotic endocarditis in 23%. Of these 41 patients, 76% had arterial emboli during life and 29% were found to have either simultaneously or sequentially the triad of venous

thrombosis, bleeding, and arterial emboli. Sack et al stressed the need for sequential monitoring of coagulation values, and the striking decreases in fibrinogen and platelet levels associated with acute vascular events. They felt this was evidence of shifts between compensated and decompensated states.

ORGAN DYSFUNCTION

Patients with DIC often have signs and symptoms of dysfunction of multiple organs. It is difficult to decide whether the dysfunctions are related to DIC or to the underlying diseases either during life or at necropsy. In literature case reports, the organ dysfunction appears to be primarily related to the underlying disease processes. Often when these processes are promptly treated, the signs of organ dysfunction and hypotension improve (7, 8, 10). The situations where organ dysfunction appears related to DIC or is out of proportion to other features of the illness include: the adult respiratory distress syndrome or shock lung (96); renal insufficiency and oliguria associated with gram-negative septicemia (97); myocardial infarction; hemolytic uremic syndrome; severe glomerulonephritis (98); amniotic fluid emboli; fresh-water drowning; cardiogenic shock syndrome with great-vessel surgery; neurologic syndromes associated with intracranial bleeding or thrombosis (5, 7, 8, 10, 99); pulmonary hemorrhage syndrome (100); rhabdomyolysis after severe exercise; amphetamines; and the infarcted skin syndrome of purpura fulminans (10, 101).

PATHOLOGY

The spectrum of pathological findings in DIC includes fibrin thrombi, major-vessel thrombosis, small-vessel thrombosis, hemorrhage, and complications thereof (100–104). The pathological hallmark of DIC is the fibrin thrombi. Fibrin thrombi are observed more frequently in the kidney (68%) than in any other organ and involve 2–10% of the glomeruli (101). Because of its rich microvasculature and high blood flow, the kidney thus acts as an excellent sieve. In contrast, fetuses have a large blood flow from the placenta through the liver, and in the stillborn infant the liver is the organ most commonly affected by the thrombi. The number of thrombi does not correlate with the duration or severity of DIC nor with the magnitude of the fibrinolytic response.

The pathology of skin lesions found in DIC, including petechiae, purpura, purpura fulminans, gangrene, acral cyanosis, and hemorrhagic bullae, has been reviewed in detail (102). Since petechiae and purpura are com-

monly associated with hematological diseases, their appearance in DIC usually leads to appropriate hematological evaluation. However, necrotic skin lesions, acral cyanosis, and hemorrhagic bullae have previously been ascribed to vasculitis, embolic disease, allergic eruptions, or poor cardiovascular states; only recently has their association with DIC been noted. The lack of inflammation of the vessel wall and the lack of involvement of the epidermis make it possible to distinguish DIC from other lesions causing palpable purpura, hemorrhagic bullae, or acral cyanoses.

Large arteries and veins are involved in many (41%) cases of DIC (101). In part, this is related to increasing use of vascular catheters for diagnostic and therapeutic maneuvers.

DIC contributes strongly to patient mortality, usually in the form of bleeding or thrombosis, especially in the lung (100), central nervous system, or gastrointestinal tract. The assignment of DIC-induced thrombosis as a major cause of death is rarely straightforward. The clearest example is the diffuse eschar secondary to thrombosis of dermal vessels (purpura fulminans). Thrombosis or embolic occlusion of the central nervous system and pulmonary vessels occurs in patients with DIC, but in nearly all instances it is associated with fungal infection, endocarditis, or intravascular catheters.

LIVER DISEASE IN DIC

Verstraete et al (105) summarized data about DIC in patients with liver disease in the *Annual Review of Medicine* in 1974. Since then, few conceptual advances have been made. Criteria for the diagnosis of DIC in the presence of liver disease have been suggested (9) and the importance of superimposed factors such as septicemia emphasized. The reduced anti-thrombin III level in liver disease (106) has been confirmed and the implications considered. Finally, a comprehensive review of diffuse intravascular coagulation in liver disease appeared in 1977 (107) emphasizing other possible mechanisms of hypofibrinogenemia and elevated FDP in liver disease that could simulate DIC. The authors emphasized caution in the use of heparin.

MANAGEMENT OF DIC

There is universal agreement that identification and treatment of all precipitating factors is the keystone of the management of DIC (1-8, 10-16, 108). Not stressed in the recent literature is the need to correct other deficits in the coagulation system or hemostasis.

Coagulation Factors

The repletion of coagulation factors is an area of controversy (1, 4). Vigorous use of plasma coagulation factors, and platelets in combination with heparin, has not been widely advocated. But Hattersley & Kunkel (109) noted that three patients treated with platelets, cryoprecipitates, and whole blood after heparinization exhibited a prompt (within 24 hours) rise in number of platelets, and their fibrinogen levels rose to almost exactly what would be predicted by the fibrinogen content of the administered material and each patient's plasma volume. These results were associated with cessation of bleeding as well. In acute promyelocytic leukemia induction therapy, heparin plus platelet transfusion is the more common treatment (71, 110).

Heparin

The use of heparin in the treatment of DIC remains controversial (1-5, 8, 16, 111) and opinions range from use of heparin in every case of DIC with laboratory abnormalities to complete avoidance of heparin. Heparin is used more frequently in DIC with thrombotic, thromboembolic, or necrotizing complications (e.g. purpura fulminans) and in the chronic DIC of malignancy than in acute hemorrhagic DIC (4, 5). A further complication is the recently discovered heterogeneity of heparin within and between batches (112).

The following clinical facts about heparin administration in DIC can be derived from the recent literature:

1. Heparin has been given to a large number of patients with laboratory and clinical evidence of DIC with relatively little morbidity or mortality.
2. The dose, schedule, and duration of treatment with heparin range from 300 to 600 U/kg per 24 hours.
3. In a review of 43 bleeding patients with acute DIC, 18 of 27 (66%) heparinized and 6 of 16 (37%) nonheparinized patients had improvement in their bleeding.
4. In the recent literature 22 of 32 heparinized (69%) and 11 of 27 (41%) nonheparinized patients survived the episode of DIC. In a review of septicemic DIC cases in nonrandomized studies we found a statistically significant survival improvement of heparinized patients (67% survival) vs nonheparinized patients (32% survival) when patients were selected who lived more than 24 hours after initiating therapy.
5. In a review, Sack et al (5) found that 29 of 48 patients treated with heparin but not antineoplastic therapy had cessation of DIC-related symptoms, and, in 19 patients, symptoms recurred with cessation of heparin. Warfarin appeared not to be effective and 12 patients nonresponsive to warfarin exhibited a heparin response.

6. There are reported episodes of DIC developing during adequate heparinization for thromboembolic problems (113), which indicates that in some instances DIC may occur by mechanisms resistant to heparinization.
7. Finally, in most cases reported in the literature, because of concomitant treatment of the underlying disease, it is difficult to know whether heparin or the response of the underlying disease is responsible for the improvement in laboratory and clinical symptoms. This may be true even of neoplasms. It would be of great value to correlate effective and objective tumor responses with cessation of symptoms of DIC. In nonrandomized studies pooled from the literature there appears to be consistent benefit for both control of bleeding and survival in heparinized patients. Equally clear are the large number of patients with DIC that do perfectly well with treatment of the underlying conditions alone. Because of the controversy over the use of heparin, a randomized trial of heparin in clinically significant DIC with both thrombotic and hemorrhagic complications appears warranted, despite all the difficulties in conducting such a trial.

A randomized trial of continuous infusion vs intermittent pulses of heparin (114) for thromboembolic disease revealed the two methods to be similar in preventing the development of pulmonary embolism, but the continuous method had fewer bleeding complications. This suggests that the continuous infusion method should be used in those patients with DIC selected to be treated with heparin. Heparin elimination and anticoagulant half-life have been studied in patients with liver disease (115) and renal insufficiency (116, 117), and the prolonged heparin retention seen in these conditions suggests that care be used in administering heparin in these clinical situations.

Other Therapies

A variety of other therapies including antiplatelet drugs (aspirin, dipyridamole, sulfinpyrazone) and fibrinolytic inhibitors and activators (epsilon aminocaproic acid, tranexamic acid, aprotinin, streptokinase, urokinase) have been suggested, but their usefulness in DIC treatment remains to be demonstrated. All textbooks continue to emphasize the contraindication of using EACA alone in DIC, and none of the recent case reports mentions EACA given with heparin.

Further advances in the management of DIC depend on the results of prospective, randomized, controlled trials of alternate therapeutic regimens, which may be conducted in the future.

Literature Cited

1. Sharp, A. A. 1977. *Br. Med. Bull.* 33:265-72
2. Bowie, E. J. W., Owen, C. A. 1977. *Semin. Hematol.* 14:341-64
3. Lerner, R. G. 1976. *Med. Clin. North Am.* 60:871-80
4. Kazmier, F. J., Bowie, E. J. W., Hagedorn, A. B., Owen, C. A. 1974. *Mayo Clin. Proc.* 49:665-72
5. Sack, G. H. Jr., Levin J., Bell, W. R. 1977. *Medicine* 56:1-37
6. Siegal, T., Seligsohn, U., Aghai, E., Kodan, M. 1978. *Thromb. Hemostasis* 39:122-34
7. Matsuda, T., Shimada, D. 1977. *Jpn. Circ. J.* 41:376
8. Colman, R. W., Robboy, S. J., Minna, J. D. 1972. *Am. J. Med.* 52:679
9. Minna, J. D., Robboy, S. J., Colman, R. W. 1974. *Disseminated Intravascular Coagulation in Man*. Springfield, Ill: Thomas. 207 pp.
10. Owen, C. A., Bowie, E. J. W. 1974. *Mayo Clin. Proc.* 49:673-79
11. Reagan, D. K., Lackner, H. 1973. *Am. J. Med. Sci.* 266:85-91
12. Yoshikawa, T., Tanaka, K. R., Guze, L. B. 1971. *Medicine* 50:237-58
13. Merskey, C. 1973. *Blood* 41:599-603
14. Marcus, A. J. 1972. *Am. J. Med. Sci.* 264:365-66
15. Corrigan, J. J. 1974. *Controversies in Internal Medicine*, ed. F. Ingelfinger, R. Ebert, M. Finland, A. Reiman, Vol. 2, pp. 623-32. Philadelphia: Saunders
16. Goodnight, S. H. 1974. *NY Acad. Sci.* 230:271-88
17. Colman, R. W., Morris, R. E., Osbahr, A. J. 1967. *Nature* 215:292
18. Bayley, T., Clements, J. A., Osbahr, A. J. 1967. *Circ. Res.* 21:469
19. Gormsen, J., Fletcher, A. P., Alkjaersig, N., Sherry, S. 1967. *Arch. Biochem. Biophys.* 120:654
20. Reimers, H. J., Kinbough-Rathbone, R. L., Cazenave, J. P., Senji, A. F., Hirsh, J., Parkham, M. A., Muskard, J. F. 1976. *Blood* 35:151-66
21. Webber, A. J., Johnson, S. A. 1970. *Am. J. Pathol.* 60:19
22. Colman, R. W. 1969. *Biochemistry* 8:1438
23. Wessler, S. 1955. *J. Clin. Invest.* 34: 647
24. Lahnborg, G., Berghem, L., Lagergren, H., Schildt, B. 1976. *Thromb. Res.* 9:653-56
25. Emeis, J. J., Lindemen, J. J. 1976. *Hemostasis* 5:193-210
26. Colman, R. W. 1969. *Biochem. Biophys. Res. Commun.* 35:272-79
27. Marder, V. J., Shulman, N. R. 1969. *J. Biol. Chem.* 244:2120
28. Marder, V. J., Shulman, N. R., Carroll, W. R. 1969. *J. Biol. Chem.* 244: 2111
29. Ratnoff, O. D. 1969. *N. Engl. J. Med.* 280:1124
30. Regan, D. H., Lackner H. 1973. *Am. J. Med. Sci.* 266:84-91
31. Ellman, L., Carvalho, A., Colman, R. W. 1973. *N. Engl. J. Med.* 288:633
32. Carvalho, A. C., Ellman, L. L., Colman, R. W. 1974. *Am. J. Clin. Pathol.* 62:107-12
33. Thomas, D. P., Niewiarowski, S., Myers, A. R., Block, K. J., Colman, R. W. 1970. *N. Engl. J. Med.* 283:663
34. Kickler, T., Cell, W. R. 1977. *Am. J. Clin. Pathol.* 68:78-80
35. Egan, E. L., Bowie, E. J. W., Kazmier, F. J., Gilchrist, G., Woods, J. W., Owen, C. A. 1974. *Mayo Clin. Proc.* 49:658-64
36. Niewiarowski, S. 1973. *Symp. Fibrinolysis Defibrination*, pp. 67-84. Paris: Librairie Arnette
37. Kirulf, P. 1973. *Scand. J. Clin. Lab. Invest.* 31:43
38. Fletcher, A. P., Alkjaersig, N., O'Brien, J., Televiski, V. G. 1970. *Trans. Assoc. Am. Physiol.* 83:159
39. Nossel, H. L., Younger, L. R., Wilner, G. B. 1971. *Proc. Natl. Acad. Sci. USA* 68:2350
40. Nossel, H. L. 1976. *N. Engl. J. Med.* 295:428-32
41. Dams, P. S., Wallare, G. A. 1975. *Thromb. Res.* 6:27-38
42. Mason, J. W., Kleeberg, U., Dolan, P., Colman, R. W. 1970. *Ann. Intern. Med.* 73:545
43. O'Donnell, T. F. Jr., Clowes, G. H. A., Talamo, R. C., Colman, R. W. 1976. *Surg. Gynecol. Obstet.* 143:539-45
44. Mason, J. W., Colman, R. W. 1971. *Thromb. Diath. Haemorrh.* 26:325
45. Tomar, R. H., Kolchins, D. 1972. *Thromb. Diath. Haemorrh.* 27:389-95
46. Branson, H. E., Wyatt, L. L., Schmer, G. 1976. *Am. J. Clin. Pathol.* 66:967-75
47. Graybill, J. R., Hawiger, J., Des Prez, R. M. 1973. *South. Med. J.* 66:410-13
48. Srichaikul, T., Puwasatier, P., Karnjanajetanee, J., Bokisch, V. 1975. *Lancet* 1:770-72
49. Kalowske, S., Howers, E. L. Jr., Margaretten, W., McKay, D. G. 1975. *Am. J. Pathol.* 78:523-36
50. Ulevitch, R. J., Cochrane, C. G., Henson, P. G., Morrison, D. C., Doe, W. F. 1975. *J. Exp. Med.* 142:1570-90

52. Muller-Berghaus, G., Lohmann, E. 1974. *Br. J. Haematol.* 28:403
53. Bohn, E., Muller-Berghaus, G. 1976. *Am. J. Pathol.* 84:239-58
54. Kramer, W., Muller-Berghaus G. 1977. *Thromb. Res.* 10:47-70
55. Lipinski, B., Gurewich, B. 1976. *Thromb. Res.* 8:403-11
56. Muller-Berghaus, G., Eckhardt, T. 1975. *Blood* 45:631-41
57. Niemetz, J. 1972. *J. Clin. Invest.* 51:307
58. Spector, J. I., Crosby, J. I. 1975. *J. Appl. Physiol.* 38:195-98
59. Ulutin, O. N., Ulutin, S. B. 1973. *N. Istanbul Contrib. Clin. Sci.* 10:254
60. Pareti, F. I., Capitanio, A., Mannucci, P. M. 1976. *Blood* 48:511-15
61. Reimers, H. J., Packham, M. A., Kinglough-Rathbone, R. L., Mustard, F. 1973. *Br. J. Haematol.* 25:675
62. Hawiger, J. 1976. *Mater. Med. Pol.* 8:206-12
63. Sharma, D. B., Bhargavam, V., Berry, A. M., Ghosh, S. 1972. *Indian J. Pediatr.* 39:375-78
64. Srichaikus, T., Nimmanitaya, S., Atrcharit, N., Siriasawakul, T., Sungpeuk, P. 1977. *Am. J. Trop. Med.* 26:525-32
65. Blix, S., Aas, K. 1961. *Acta Med. Scand.* 169:63
66. Shanberge, J. N., Tanaka, K., Gruhl, M. C. 1971. *Am. J. Clin. Pathol.* 65:723
67. Megaffin, B. B., Crow, E. W. 1977. *J. Kans. Med. Soc.* 78:374-78
68. Siebert, W. T., Natelson, E. A. 1976. *Arch. Surg.* 111:539-41
69. Leavy, R. A., Kahn, S. B., Brodsky, I. 1970. *Cancer* 26:142-45
70. Samaha, R. J., Bruns, T. M. C., Ross, G. J. 1973. *Arch. Surg.* 106:295-98
71. Gralnick, H. R., Marchesi, S., Givelber, H. 1972. *Blood* 40:709-18
72. Mintz, U., Djaldetti, M., Rosenszajn, L., Pinkhas, J., DeVries, A. 1973. *Biomedicine* 19:426-30
73. German, H. L., Smith, J. A., Lindenbaum, J. 1976. *Am. J. Med.* 61:547-52
74. Lee, C. A., Moss, S. 1976. *Proc. R. Soc. Med.* 69:521-22
75. Weiss, H. J., Phillips, L. J., Hopewell, W., Christy, N. P., Mitty, J. F. 1973. *Am. J. Med.* 54:653
76. Hasiba, U., Rosenbach, L. M., Rockwell, D., Lewis, J. N. 1975. *N. Engl. J. Med.* 292:505-7
77. Bonilla, C. A. 1975. *Thromb. Res.* 6:151
78. Schmaier, A., Colman, R. W., Claypool, W. 1978. *Circulation* 58:II-215
79. Druskin, M. S., Drijanski, R. 1972. *J. Am. Med. Assoc.* 219:755-56
80. Goodnight, S. H., Kenoyer, G., Rapaport, S. I. 1974. *N. Engl. J. Med.* 290:1043-47
81. Drayer, B. P., Posner, C. M. 1975. *J. Am. Med. Assoc.* 219:174-75
82. Beller, F. K., Uszynski, M. 1974. *Clin. Obstet. Gynecol.* 17:250-78
83. Kleiner, G. J., Greston, W. M. 1974. *S. Afr. Med. J.* 49:446-52
84. MacKenzie, I. Z., Sayers, L., Bonnar, J., Hiltier, K. 1975. *Lancet* 2:1066-69
85. Hutter, J. J., Hathaway, W. E., Wayne, E. R. 1976. *J. Pediatr.* 88:1026-31
86. Patel, C. C. 1977. *Pediatr. Clin. North Am.* 24:579-84
87. Ports, T. A., Dewel, T. F. 1977. *Ann. Intern. Med.* 87:60-61
88. Blatt, P., Lundblad, R., Kingdon, H., McLean, G., Roberts, H. 1974. *Ann. Intern. Med.* 81:766
89. Triantphylopouloues, D. C. 1972. *Am. J. Clin. Pathol.* 57:603
90. Ihemarho, H. W. C., Breeze, G. R., Fletcher, D. J., Stuart, J. 1973. *Lancet* 1:231-34
91. Mehta, B., Briggs, D. K., Sommers, S. C., Karparkin, M. 1972. *Am. J. Med. Sci.* 264:353-63
92. Umlas, J. 1976. *Transfusion* 16:640-43
93. Boyd, A. D., Engelman, R. M., Beacdet, R. L., Larkner, H. 1972. *J. Thorac. Cardiovasc. Surg.* 69:685-93
94. Krug, H., Raszeja-Wanic, B., Wochowiak, A. 1974. *Ann. Intern. Med.* 81:494-97
95. Colman, R. W., Minna, J. D., Robboy, S. J. 1977. *Int. J. Dermatol.* 16:47
96. Bone, R. C., Francis, P. B., Pierce, A. K. 1976. *Am. J. Med.* 61:585-89
97. Preston, F. E., Malia, R. G., Sworn, M. J., Blackburn, E. K. 1973. *J. Clin. Pathol.* 26:120-25
98. Robson, A. M., Cole, B. R., Kienstra, R. A., Kissane, J. M., Alkjaersig, N., Fletcher, A. P. 1977. *J. Pediatr.* 90:881-92
99. Collins, R. C., Al-Mondhiry, H., Chernik, N. L., Posner, J. B. 1975. *Neurology* 25:795-806
100. Robboy, S. J., Minna, J. D., Colman, R. W., Birdorf, N. I., Lopas, H. 1973. *Chest* 63:718
101. Robboy, S. J., Colman, R. W., Minna, J. D. 1972. *Hum. Pathol.* 3:327
102. Robboy, S. J., Mihm, M. C., Colman, R. W., Minna, J. D. 1973. *Br. J. Dermatol.* 88:221
103. Robboy, S. J., Salisbury, K., Ragsdale, B., Bobroff, L. M., Jacobson, B. M., Colman, R. W. 1971. *Arch. Intern. Med.* 128:790
104. Stewart, G. J. 1977. *Thromb. Haemostasis* 38:831

105. Verstraete, M., Vermeylen, J., Collen, D. 1974. *Ann. Rev. Med.* 25:447-55
106. Braunstein, K. M., Eurenis, K. 1976. *Am. J. Clin. Pathol.* 66:488-94
107. Straub, P. W. 1977. *Semin. Thromb. Hemostasis* 4:29-39
108. Heene, D. L. 1977. *Semin. Thromb. Hemostasis* 3:291
109. Hattersley, P. G., Kunkel, M. 1976. *Transfusion* 16:641-48
110. Drapkin, R. L., Gee, T. S., Dowling, M. D., Arlin, Z., McKensie, S., Kempin, S., Clarkson, B. 1978. *Cancer* 41:2484
111. Straub, P. W. 1974. *Thromb. Diath. Haemorrh.* 33:107-12
112. Rosenberg, R. D. 1977. *Semin. Hematol.* 14:427-40
113. Klein, H. G., Bell, W. R. 1974. *Ann. Intern. Med.* 80:477-81
114. Salzman, E. W., Deykin, D., Shapiro, R. M., Rosenberg, R. 1975. *N. Engl. J. Med.* 292:1046
115. Tien, A. N. 1977. *Thromb. Haemostasis* 38:701
116. Perry, P. J., Herron, G. R., King, J. C. 1974. *Clin. Pharmacol. Ther.* 16: 514-19
117. Tien, A. N., Bjornson, J. 1976. *Scand. J. Haematol.* 17:29-35

IMMUNOBIOLOGY OF THE MATERNAL-FETAL RELATIONSHIP

♦7330

*Ross E. Rocklin, M.D.*¹

Department of Medicine, Harvard Medical School, Robert Breck Brigham
Hospital, Boston, Massachusetts 02120

John L. Kitzmiller, M.D.

Department of Obstetrics and Gynecology, Harvard Medical School,
Boston Hospital for Women, Boston, Massachusetts 02115

Michael D. Kaye, Ph.D.

Department of Obstetrics and Gynecology, Sydney University,
Sydney, Australia

INTRODUCTION

The fetuses of all outbred mammalian species present an immunogenetic set of paternal antigens that are foreign to the mother. Yet the fetus will survive the period of intrauterine gestation without rejection despite the fact that its mother will reject fetal or paternal tissues grafted at any other site. This apparent paradox of transplantation immunology still remains an enigma today even though it has been the focus of active research over the past twenty years. Medawar (1) hypothesized many years ago that the success of the fetus as an allograft could be explained as follows: (*a*) the conceptus was not immunogenic and therefore did not invoke a maternal immunologic response; (*b*) pregnancy altered the maternal immunologic response; (*c*) the uterus was an immunologically privileged site; (*d*) the placenta was an effective immunologic barrier; and (*e*) the fetus was not yet immunologically competent. While many of these hypotheses did not withstand the test of time, they provided investigators with the opportunity to critically analyze, and as a result enhance understanding of, this very complex phenome-

¹Present address: Department of Medicine, Box 30, T-NEMC, 171 Harrison Avenue, Boston, Massachusetts 02111.

non. In this chapter we review the immunobiology of the maternal-fetal relationship, including the factors that lead to maternal sensitization, the maternal responses to fetal antigens, the factors that alter these responses, the fetal responses to maternal antigens, and a summary of the mechanisms that might allow fetal survival.

MATERNAL EXPOSURE AND LOCAL RESPONSE TO ANTIGENS RELATED TO CONCEPTION

Antigens of Semen and Sperm: Local and Systemic Responses

Semen contains many antigenic components from the testis and accessory glands of the male reproductive tract as well as spermatozoa (2). There is haploid expression of the major histocompatibility antigens on spermatozoa in man (3) and the mouse (4). An enzyme, lactic dehydrogenase-X (LDH-X), appears to be specific for spermatozoa (5), while the carbohydrate ABO antigens on spermatozoa are probably the result of adsorption rather than haploid expression (4). Seminal plasma or "sperm-coating antigens" are present in the acrosomes of ejaculated spermatozoa, and their removal may have something to do with the process of capacitation within the female genital tract (2).

Female immune responses to semen and sperm can be arbitrarily divided into antibody production and lymphocyte reactivity. The plasma cells of the cervix appear to be the major producers of immunoglobulin (Ig) in the female genital tract. Cervical mucus contains IgA and SIgA (two serum IgA molecules plus a secretory component), and IgG and small amounts of IgM and complement factors. Mucus also exhibits antibacterial and phagocytic activities caused by lysozymes and opsonizing antibodies (2). Sperm recovered from the uterus in the mouse and rabbit have antibodies attached, while those from the oviduct do not (6). This suggests that sperm populations may differ in their immunologic interaction with the female genital tract. That "selection" by the female genital tract occurs is amply demonstrated by the observation that only 5% of sperm migrate upwards from the vagina in mammals. In the uterus, most spermatozoa undergo phagocytosis by leukocytes within a few hours. The concept that spermatozoal selection may be the result of errors in spermatogenesis (7) has been refuted (8).

The introduction of allogeneic washed epididymis spermatozoa into the uterus of rats sensitizes the animals to later intradermal injection of lymphoid cells of the same strain, and also produces a marked local inflammatory response 24-48 hours after reexposure to the spermatozoa. Subsequent mating with allogeneic males enhances the reproductive efficiency of the sensitized uterine horn. Inoculation of virgins with allogeneic sperm pro-

duces enlargement of the uterine regional nodes when compared with inoculation by syngeneic sperm. The results suggest a local cellular response to spermatozoa within the uterus (9).

Infertility due to humoral factors has been ascribed to women with otherwise unexplained inability to conceive, is controversial, and has been reviewed elsewhere (2).

The Blastocyst

The ovum is surrounded by a mucoprotein coat, the zona pellucida. Contact of the sperm with the zona pellucida in the fallopian tube is species specific and depends on the interaction of terminal sugars on the cell surface glycoproteins of both (10). Once fertilization occurs, the zona pellucida becomes impenetrable to further spermatozoa because of the active mucoids or glycoproteins released from the cortical granules and diffused through the perivitelline space to the zona (11). The zona may protect the blastocyst during passage through tube and uterus, although zona-free eggs will survive and implant in the uterus of sensitized recipients (12). In the mouse, immunoglobulins coat the blastocyst by the time of implantation (13) and the zona pellucida is permeable to them (14). The source and biological role of these immunoglobulins are unknown.

There is ample evidence that preimplantation mouse blastocysts are immunogenic. They will be rejected upon transfer to ectopic sites in presensitized hosts (15), although the presence of the zona pellucida prolongs time to rejection (16). As an initial challenge in other recipients, blastocysts develop into trophoblast and induce specific unresponsiveness to later skin grafts from the donor strain (17). Trophoblast cells grown in culture from four-day mouse blastocysts somehow inhibit the *in vitro* growth of macrophage monolayers (18). Minor histocompatibility antigens (non-H-2) have been demonstrated on the trophectoderm surface of the 4–8-cell zygote, the morula, and the early murine blastocyst by immunoperoxidase labeling in the mouse (16). Other workers confirmed these results by cytotoxicity and immunofluorescence techniques (17, 19–21). Cleavage eggs and the inner cell mass also express tumor-related embryonic antigens (teratoma-like; SV-40), which may be lost at 4½ days. Most investigators could not identify major H-2 antigens on the outer trophectoderm of the blastocyst, but they are present on the inner cell mass (19, 22, 23), which differentiates to embryo, yolk sac, and amnion.

The Placenta

IMPLANTATION As implantation nears, the zona pellucida is lost, and the outer trophectoderm seems to lose its non-H-2 antigenic determinants along with a decrease in concanavalin A and colloidal iron binding (22, 24). It is interesting that this occurs at the time of first intimate contact with

maternal tissues, and in conjunction with the development of polyploidy of the outer trophoblast cells and the appearance of their phagocytic and secretory properties. Trophoblast differentiation *in vitro* also leads to loss of surface antigens, so it is not dependent on the maternal environment (22).

At implantation in the mouse, trophoctoderm penetrates the uterine epithelium and further differentiates mural trophoblast to primary giant cells, polar trophoblast to secondary giant cells and ectoplacental cone. Trophoblast invasion of decidua probably occurs through a combination of cytoplasmic penetration around maternal cells, phagocytosis, and diffusion of cytolytic enzymes. In hemochorial placentation common to primates and laboratory rodents, maternal capillaries are tapped and the trophoblast villi or labyrinths are bathed with maternal blood. Cytotrophoblast migrates up uterine arteries to or beyond the myometrium, which induces dramatic structural changes that lead to expansion of the vessels. Giant cells also migrate through decidua and myometrium.

In primates the proliferative core cells (cytotrophoblast) undergo cytoplasmic differentiation, and the cell membranes break down to form a syncytium, the outer syncytiotrophoblast. The microvilli and glycocalyx of this layer of trophoblast are in contact with maternal blood.

Contact of the blastocyst with the uterine wall begins the process of decidualization whereby the uterine and stromal cells undergo rapid DNA synthesis and mitosis (25). In the mouse the decidual cells increase their DNA content by endoreduplication (26), and the trophoblast cells cease dividing as they move away from the influence of the inner cell mass (27) but continue their DNA synthesis by endoreduplication (28).

Whether or not there is cell fusion between trophoblast and uterine cells is controversial. The process was first noted in the Australian marsupial Native cat (29) and bandicoot (30, 31). This was confirmed in woman (32, 33) by light microscopy, and at the electron microscope level in the bandicoot (34), rabbit (35, 36), and hamster (37). Other investigators noted fine decidual cytoplasmic protrusions approaching human trophoblast cells, but no evidence of cell fusion (38). The trophoblast cells are phagocytic, and they will incorporate labeled maternal nuclear material from the uterus of rats (39) or mice (40). Decidual cells similarly incorporate label from the nuclei of mouse blastocysts labeled *in vitro* with ^3H -thymidine prior to their transfer to pseudopregnant recipients (41). These observations may have important implications in the development of a barrier between fetal and maternal tissues. When the cells of different species hybridize *in vitro* and nuclear mixing occurs, the resultant hybrid cell undergoes a marked alteration in its cell surface antigens (42, 43). Other investigators believe the high DNA count of giant cells arises from endoreduplication rather than from cell fusion (27). Studies of artificially implanted embryos did not reveal

hybrid enzymes, which would be expected if maternal cells fused with trophoblast (28, 44).

PLACENTAL ANTIGENS A variety of methods have been employed to determine whether postimplantation trophoblast is immunogenic. Since some of these methods disrupt the normal anatomical arrangements of the placental tissue, one should bear in mind that conclusions about trophoblast antigens may be artifactual. Extracts of mouse and rat placentae have been used to raise antisera that will then cause placental damage and abortions in pregnant animals of the same strain (45, 46). The antisera are absorbed with somatic cells or tissues prior to injection, which suggests that there are organ-specific antigens in the placenta. These techniques, however, cannot determine which type of placental cell is contributing the antigens. Youtananukorn et al (47) applied the papain digest method of preparing soluble, surface histocompatibility antigens to the human placenta, and determined that autologous lymphocytes were sensitized (migration inhibitory factor production) to the "placental antigens." Finally, placental fraction with enzyme specificities of plasma membrane exhibited only 5% of the HLA-A and B and β_2 -microglobulin titers found in spleen cells (48).

Transplantation of murine ectoplacental cone trophoblast to ectopic sites revealed that trophoblast growth is not affected by prior sensitization with donor-strain skin grafts or injections of spleen cells (49, 50). The conclusion that implanted, differentiated trophoblast is not immunogenic is supported by more recent studies in which antibody binding to H-2 and non-H-2 determinants on ectoplacental cone derivatives was negative, while antibody binding to embryonic sacs was strongly positive (21, 22). Beer et al (51) showed, however, that rat ectoplacental cone trophoblast transplanted to renal capsules of allogeneic females (*a*) induces hypertrophy of draining lymph nodes, (*b*) induces production of paternal-strain hemagglutinins, and (*c*) slightly impairs subsequent alloimmune responses of the hosts.

Since there is a highly sulfated sialomucoprotein substance on the surface of murine trophoblast cells (52), many have suggested that this coat may render trophoblast nonimmunogenic by masking antigenic sites or repelling lymphocytes by the high electronegative charge. Currie et al (53) reported that treatment with neuraminidase to remove sialic residues enabled ectoplacental cone trophoblast to induce transplantation reactions, but this result was not confirmed by Simmons et al (54). Billington (17) cautions that experiments with explants of ectoplacental cones may be misleading, since all of the trophoblasts become multinucleate, polyploid cells.

Immunologic studies of trophoblast cells in tissue culture are fraught with similar difficulties concerning the origin and nature of the reacting cells. Anti-HLA sera and complement were cytotoxic to human placental

cells from early pregnancy, although the density of HLA antigenic sites was much less than on fetal skin cells (55). Two groups of investigators observed that autologous leukocytes (but not purified lymphocytes) were cytotoxic to trypsinized human placental cells in culture, and that trophoblast cells induced a blastogenic response in the lymphocytes (56, 57). The leukocytotoxic effects on the placental cells were blocked by autologous serum (57). Using the immunoperoxidase-labeling technique, Jenkinson & Billington (58) showed that early embryonic cells grown in culture were certainly positive for histocompatibility antigens as compared to placental cells. However, Billington and co-workers (21) most recently studied postimplantation trophoblast in culture with a mixed hemadsorption technique, and they found that *some* trophoblast cells were positive for non-H-2 antigens. They present a cogent discussion of possible reasons for trophoblast cells to shed antigens after implantation, and point out that inner diploid cells may remain antigen positive, while outer differentiated polyploid cells are negative (17, 21).

Frozen sections of intact human placentae have been studied by immunohistologic techniques, and many immunoproteins have been identified (59). In late pregnancy, IgG is primarily localized in the trophoblast basement membrane, while IgM and C3 are found in deposits of intervillous fibrin (60-63). HLA-reactive sites and β_2 -microglobulin were present in mesenchymal and fetal endothelial cells, but not on the surface or basement membrane of trophoblast cells (64). In conclusion, it is perhaps useful to recall Billington's statement that a weakly antigenic trophoblast, despite its failure to induce a state of transplantation immunity in the host, still may be responsible for the induction of circulating antibodies (17).

THE UTERUS AS AN IMMUNOLOGICALLY PRIVILEGED SITE A question has been raised concerning whether or not the uterus is an immunologically privileged site similar to the anterior chamber of the eye and the hamster cheek pouch, where poor lymph drainage is related to failure of graft rejection (65). This would imply that the reason the conceptus thrives within the uterus is that either the afferent or efferent arms (or both) of the expected allograft response are ineffective at that site. The concept does not explain, of course, the rare human abdominal pregnancy that proceeds to term. The idea of a protective intrauterine environment is suggested by experiments in which blastocysts transplanted to ectopic sites in presensitized mice are destroyed, while blastocysts transferred to the uterus develop normally (49).

Beer & Billingham (9) carried out an elegant series of experiments demonstrating that the decidual response produced by pseudopregnancy or hormone injections leads to prolongation of a graft of allogeneic skin within

the uterus, but that decidua were not enough to prevent rapid rejection of intrauterine skin grafts in presensitized hosts (9). The possible role of decidua in limiting recognition of the conceptus as an allograft remains unclear, since there is lymphatic drainage of the decidua and myometrium, and lymphocytic infiltration near the trophoblast is observed in human pregnancy (66, 67). Recent ultramicroscopic studies demonstrate intermingling of human decidual and trophoblastic cells without structural elements as a barrier between them (38, 68). Fibrillar deposits were distributed randomly around the deciduotrophoblastic complexes, and there is some evidence this material is produced by the maternal decidual cells (38, 68). A close association of small developing decidual cells with lymphoblasts around precapillary arterioles was also noted (38).

A variety of local immune responses can be induced within the uterus of laboratory animals. Transfer of allogeneic skin, leukocytes, or lymph node cells to one uterine horn stimulates ipsilateral hypertrophy of the para-aortic lymph nodes, sensitizes the host for accelerated rejection of a subsequent skin graft, and *locally* sensitizes the uterus for a delayed inflammatory reaction to a second inoculum of lymphoid cells (9). Thus, the uterus does not seem to be protected from participation in immune reactions. Interestingly, allogeneic trophoblast also induces hypertrophy of the para-aortic nodes in rats, mice, and hamsters (9), while in contrast a decrease in size and cellularity of germinal centers of the pelvic and para-aortic lymph nodes in human mid- and late pregnancy has been described (69, 70). More recent experiments in rats indicate that removal of para-aortic nodes prior to allogeneic pregnancy preserves maternal responses to paternal antigens subsequent to pregnancy, which are otherwise impaired after control pregnancies (51). Adoptive transfer of para-aortic node cells (or serum) from multiparas suppresses host antibody response to paternal-strain skin grafts. Based on these experiments and others (71), Beer suggests that lymph nodes draining the uterus may be the site of generation of immunosuppressive T cells or blocking antibody (see below).

THE PLACENTA AS AN IMMUNOLOGIC BARRIER Undoubtedly, a major factor in the immunological coexistence of gravida and fetus is their ordinarily complete vascular separation. Interposed is the placenta, which conceivably could prevent immunogenic fetal cells from reaching the maternal circulation or protect the fetus from maternal lymphocyte attack. However, the placenta is an incomplete barrier in this respect, and it is well known that small amounts of fetal blood containing red cells, leukocytes, and lymphoid cells enter the maternal circulation prior to termination of pregnancy (72-75). Occasionally, an Rh-negative gravida will be sensitized to Rh-positive fetal red cells during her first pregnancy, although this is

much more likely following the larger fetal-to-maternal transfusion associated with labor and delivery. The passage of maternal lymphocytes to the fetus is apparently a much rarer event, but is not unheard of (73). Besides cellular material (including deportation of syncytiotrophoblast fragments to maternal lungs), placental glycoprotein hormones, which are at least immunogenic across species (xenogeneic), also pass to the maternal circulation. However, in primates the cross reactivity of human chorionic gonadotropin (hCG) with maternal gonadotropins (FSH and LH) and of human placental lactogen (hPL) with growth hormone may be protective.

The placenta probably acts as a selective filter for maternal immunoglobulins reaching the fetus. Fc receptors are identified on placental cells. IgG is transported across the placenta, but not IgM or IgA. In rodents, immunoglobulin transport across the yolk sac is also important (76). Cytotoxic antibodies to fetal antigens are absent in neonatal blood (77), perhaps because of sequestration in the placenta (78, 79). Johnson, Trenchev & Faulk (80) also suggested that immune complexes might be removed from fetal blood by deposition in the placental stem vessels (80).

Embryonic Antigens and Fetal Blood Cells

As noted above, the earliest appearance of the major histocompatibility (H-2) antigens was found on cells of the inner cell mass of mouse embryos by the late blastocyst stage (96 hours after fertilization) (23) or slightly later (19, 81). As development proceeds, antigenicity increases at different rates in various embryonic organs (82).

The red cell carbohydrate antigens of the ABO system are present from 5-6 weeks of gestation but are not fully mature until 2-4 years of life. These antigens are present on all somatic cells as well as on certain intestinal gram-negative bacteria, which accounts for the presence of isoagglutinins in O-positive, six-month-old infants. These facts also account for the observation that firstborn infants may be affected by ABO incompatibility, and that increasing severity with subsequent children may not occur since the fetal red cells are poorly immunogenic and the fetal somatic cells compete with the red cells for the maternal IgG (83).

The major histocompatibility antigens are present on red cells in rats, but not in man. These antigens are present on human fetal leukocytes, and antibodies are found in the serum of 25% of multiparous women (84, 85). Maternal exposure to allogeneic fetal leukocytes may not be obligatory for the production of anti-HLA antibodies, however, since they have been detected in maternal sera after molar pregnancies (86). Women exposed to sperm do not develop anti-HLA antibodies prior to their first pregnancies (87).

MATERNAL RESPONSES TO THE FETUS

General Maternal Immunocompetence

An important issue concerning survival of the fetus is focused on the ability of the mother to mount an immunologic response to foreign (paternal) antigens. Is the mother immunocompetent in this regard throughout the period of gestation, and if not, does her inability to respond reflect a specific or more generalized defect? The answers to these questions are complex and the data incomplete.

The morphology and histology of lymphoid tissue and the enumeration of T and B cells in pregnancy has been evaluated in man as well as in animals. There is an involution of the thymus during pregnancy (69) and a decrease in the size and cellularity (B cells) of germinal centers present in para-aortic lymph nodes (70). However, the weight and proportion of T cells within the spleen and mesenteric lymph nodes is apparently increased. The overall percentage of, and total lymphocyte count is essentially unchanged. However, there are fewer circulating T cells early in pregnancy (up to 20 weeks gestation) with a compensatory increase in B cells (88, 89). This reversal of the normal T:B cell ratio is gradually corrected after 20 weeks of gestation. It is quite possible that the apparent reduction in circulating T cells in early pregnancy represents not a "loss" of these cells but rather their redistribution or transfer to another lymphoid compartment. Data quantifying the various subsets of T cells would be extremely valuable in determining whether populations of helpers or suppressors are being selectively affected.

Humoral and cellular immune responses to a variety of antigens have been measured. In most studies, the plasma concentrations of IgG, IgM, and IgA in pregnancy are normal compared to those in the nonpregnant state (90). However, IgG levels decrease somewhat late in pregnancy, possibly because of placental transfer (91). Little information is available concerning IgD and IgE levels, although the former has been reported to be elevated at term (92). The production of specific antibodies to environmental antigens is normal, which suggests that these aspects of T and B cell function in general are intact during pregnancy (93). Tuberculin skin reactivity is either normal (94) or slightly depressed (25% of cases) in the last trimester of pregnancy (95). Graft rejection, another *in vivo* manifestation of cellular immunity, is marginally impaired during pregnancy. The rejection of primary ("first set") skin grafts is abnormal, but the secondary rejection of a graft from the same donor is normal (96). Two other components of the immune system, complement, and phagocytosis and bactericidal activity, have also been evaluated during pregnancy. The total

complement level (CH_{50}) and the third component (C3) are progressively elevated during the course of pregnancy although it is not clear whether this reflects inflammatory changes or, more likely, is hormonally related (97). Leukocytes from pregnant women show markedly increased rates of phagocytosis and bactericidal activity (98). Whether the latter finding is also a result of hormonal influences remains to be clarified. Thus, during pregnancy the various components of immunity per se are basically intact with only minor alterations.

Humoral Antibody Responses

A wide range of antibodies are produced against fetal antigens. The production of anti-Rh antibody by Rhdd women carrying RhD fetuses is a well known example of this response (83). Furthermore, mothers whose red cells lack A or B red cell antigens and who carry a fetus possessing these antigens usually exhibit a rise in anti-A or anti-B titers.

The gravida also develops antibodies to paternal HLA-A, B, and C series antigens. These antibodies will be detected in 20–30% of primagravid women, and the incidence and titers increase with succeeding pregnancies (99, 100). Titers of antibody rise during pregnancy and decrease slightly after delivery. However, in some subjects these antibodies persist for months to many years after the last pregnancy. In addition, non-anti-HLA cytotoxic antibodies can be detected in maternal sera and exhibit different kinetics. These antibodies decrease during the third trimester and reappear after delivery. It is tempting to attribute the disappearance of these antibodies to the combination of antibody with fetal antigen that is shed into the maternal bloodstream.

Recently, IgG antibodies have been described that are not directed against HLA (ABC) antigens, are noncytotoxic, and are directed against HLA-D antigens (101). These antibodies are of particular interest because they identify antigens present on paternal cells that are thought to be involved in initiating the mixed lymphocyte reaction in vitro and presumably in vivo as well. That is, these antigens are important for triggering sensitized effector lymphocytes, and the presence of these antibodies in vitro blocks the reaction. The latter may be analogous to the anti-Ia antibodies described in the mouse (102). Their role in the survival of the fetus is discussed more fully in the section on specific immunosuppression.

Cellular Immune Responses

Through use of a variety of in vitro assays to measure maternal lymphocyte responses, including proliferation, cytotoxicity, and lymphokine production, it has been demonstrated that maternal cells are clearly reactive against paternal antigens. In the presence of normal serum (but not autolo-

gous serum), cells from pregnant women proliferate as well as cells from nonpregnant females when mixed together in vitro with paternal cells (103–106). The maternal response is somewhat stronger, however, against third-party cells. The reasons for this disparity will also be discussed in the section on specific immunosuppression. Furthermore, when fetal lymphocytes are used to stimulate maternal cells, a weak proliferative response is observed (107). The mechanism for the latter observation is not clear but may relate to the ability of fetal mononuclear cells to secrete soluble factors that have inhibitory effects on proliferation (108).

Cells that are directly cytotoxic (killer T cells) against fetal cells, and to a lesser extent against paternal cells as well, are also found in the maternal circulation (109). These cytotoxic lymphocytes are rarely found circulating in the blood of normal individuals, since they are usually only generated in vitro during the course of a mixed lymphocyte reaction (MLR). Therefore, the presence of these cells in the blood of pregnant women, as well as the finding that these cells exhibit cytotoxicity toward fetal/paternal but not third-party cells, suggests that an active sensitization to paternal antigens is occurring in vivo. Of further interest, maternal lymphocytes also exhibit cytotoxicity toward trophoblast tissue from their own pregnancies and, to a lesser extent, for trophoblast cells from other pregnancies (57). Since trophoblast cells bind large amounts of IgG, even after vigorous washing, this phenomenon may be mediated by lymphocytes (K cells) capable of binding IgG through their Fc receptor, which activates them for killing (antibody-dependent cellular cytotoxicity).

Further evidence of a state of maternal sensitization to paternal antigens was demonstrated in studies in which lymphokine production was measured. Lymphocytes from pregnant women cultured in vitro with cord blood or paternal mononuclear cells or placental antigens produce macrophage migration inhibitory factor (MIF) within 24 hours (110, 111). This maternal response is indicative of a state of specific presensitization since mononuclear cells from randomly mixed normal donors did not produce MIF within this period of culture, and maternal cells usually do not make MIF against third-party cells (112). In addition, there was a correlation between the presence of MIF production and the parity, which indicates that sensitization is progressive with repeated exposure to antigen (111). Taken together, these data imply that washed maternal lymphocytes, cultured in the absence of autologous serum, are perfectly capable of responding in a number of ways against cells bearing paternal antigens.

Nonspecific Immunosuppression

STEROID HORMONES The placenta synthesizes estrogens from fetal and maternal precursors and secretes them into maternal and fetal circulations

in increasing concentrations as pregnancy proceeds. Maternal serum levels of the most active estrogen, estradiol, are ~ 20 ng/ml at term, but the concentration is ~ 50 ng/ml in the intervillous blood (113).

Early studies failed to show an immunosuppressive effect of estradiol on experimental allograft rejection, but the tissue levels of estrogen may have been too low (114, 115). More recently, estradiol has been shown to prolong first- and second-set skin grafts in mice (116) and to inhibit the rejection of corneal grafts in preimmunized rabbits (117). In another study estradiol only prolonged first-set allogeneic skin grafts in sublethally irradiated mice (as compared to nonirradiated mice). The authors speculated that the estrogen impaired the development or proliferation of thymic stem cells during recovery from irradiation (118). Estradiol certainly involutes the thymus in adult experimental animals (69). Estrogen and hCG in combination produced a marked decrease in thymic weight and in the number of small lymphocytes in the thymus (118). As noted above, involution of the maternal thymus is well recognized in human pregnancy. The murine thymus is able to metabolize progesterone by 20- α -hydroxylation to a much greater degree than lung or kidney tissue (119). Isolated thymocytes have even greater specific activity, which suggests a role for progesterone in thymic function.

Progesterone is synthesized and secreted by trophoblast cells into the maternal blood, and thus is in much higher concentration in the intervillous blood (886 ng/ml) than in peripheral blood (160 ng/ml, late pregnancy) (120a). Placental concentration of progesterone is $3.1 \mu\text{g/gm}$ wet weight at term (120b). Injection of progesterone into experimental animals did not prolong allograft survival in several early studies (114, 116, 121). The dosage or route of administration may have been inadequate, since subsequent experiments showed that progesterone or a synthetic progestin extended viability of homologous skin grafts in rhesus monkeys or of heart grafts in rats, respectively (122, 123). Siiteri and colleagues (124) first studied the effects of progesterone released into local tissue in high concentrations from tubular Silastic® inserts. In these experiments, later confirmed by Beer & Billingham (125), tissue concentrations of 50–200 $\mu\text{g/ml}$ progesterone suppressed the expected inflammatory reaction to cotton threads and prolonged the survival of skin xeno- and allografts in rats.

The effects of reproductive steroids on *in vitro* models of cellular immunity have been widely studied in recent years. Estradiol, progesterone, and diethylstilbestrol in the dose range of 10–50 $\mu\text{g/ml}$ have all been shown to reduce the lymphocyte proliferative response to the mitogenic plant lectins phytohemagglutinin and concanavalin A (126–128). In human studies of the mixed lymphocyte reaction, doses as low as 0.2 $\mu\text{g/ml}$ estradiol slightly stimulated lymphocyte reactivity, while higher doses of 20 $\mu\text{g/ml}$ of es-

tradiol or progesterone inhibited the uptake of tritiated thymidine by responding lymphocytes (128, 129). Washing lymphocytes removed the inhibitory effect of the reproductive steroids (129, 130). One study showed that estradiol must be present from the beginning of the five-day mixed lymphocyte culture in order to show its inhibitory effect (129); however, Siiteri and co-workers (130) demonstrated that progesterone (20 $\mu\text{g}/\text{ml}$) could be added to the end of the five-day mixed lymphocyte culture and it still would inhibit the uptake of tritiated thymidine, which is the marker for cellular proliferation. Thus, the latter suggests that progesterone is suppressing not cell activation but incorporation of labeled thymidine into DNA. Reproductive steroids do not produce cytotoxic effects on the lymphocytes at the effective dose ranges.

We have examined another model of cellular immunity, the production of migration inhibitory factor (MIF) by presensitized guinea pig lymphocytes. Neither estradiol nor progesterone in concentrations of 2–20 $\mu\text{g}/\text{ml}$ significantly inhibited either the production of MIF or the macrophage response to preformed MIF (131).

There are fewer studies of the effects of reproductive steroids on humoral immune responses. Treatment of mice with 2.5 μg of estradiol resulted in increased numbers of splenic antibody-producing cells in mice sensitized to *Escherichia coli* (132). Incubation of murine spleen and thymic cells with low doses of estradiol (500–5000 pg/ml) produced significant increases in numbers of antibody-producing cells and in the uptake of tritiated thymidine by the lymphoid cells (132). Much higher doses of estradiol slightly suppressed these aspects of the humoral and cellular immune systems. Beer's most recent studies indicated that the lowest dose of locally released progesterone (4 ng/ml serum, 100 ng/ml tissue) caused earlier initiation of antibody production in rats stimulated with dermal allografts (125). Higher doses of progesterone (186 ng/ml serum, 200 ng/ml tissue) significantly inhibited the formation of hemagglutinating antibodies in these experiments.

The plasma levels of transcortin, transcortin- and albumin-bound cortisol, and free unbound cortisol all increase progressively during pregnancy. The increased cortisol levels (5–20 $\mu\text{g}/\text{ml}$) probably result from reduced hepatic clearance of cortisol secondary to estrogen, independent of the increase in transcortin synthesis, also an estrogen effect (133). Maternal cells and tissues are thus exposed to an average daily concentration of cortisol at least twice normal. Cortisol was early shown to delay allograft rejection in animals (122), and corticosteroids are used clinically in human renal transplantation. Corticosteroids have cytolytic effects on lymphoid cells of some species (mouse, rabbit, rat) but not of guinea pigs or primates (134). Decreased circulating lymphocyte counts produced by phar-

macologic doses of corticosteroids result from a redistribution of B and T cells (135). Low doses of cortisol may enhance lymphoid cell proliferation, but sufficiently high concentrations nonspecifically suppress DNA synthesis in basal and stimulated lymphoid cells (136). Cortisol inhibits MLC response (uptake of ^3H -thymidine) in the dose range of 1–20 $\mu\text{g}/\text{ml}$, but the effect is different from that of progesterone, in that cortisol must be present early in the five-day culture for maximum suppressive effect (130). Washing lymphocytes preincubated with cortisol does not reduce the MLC suppressive effect, as it does with progesterone (130).

Corticosteroids have pronounced effects on the macrophage. Aggregation or migration inhibition responses to lymphokines are suppressed (137, 138a), phagocytic activity is decreased, and the binding activity of membrane receptors for IgG and C3 is reduced (135). Corticosteroids may influence cell membrane function directly. It is probable that the effects of corticosteroids on allograft and delayed hypersensitivity reactions depend more on anti-inflammatory effects (decreased capillary permeability; increased vascular tone; decreased macrophage responses) than on direct suppression of lymphocyte function. Cortisol may be concentrated at the placental surface, since a transcortin-like protein with immunosuppressive effects has been demonstrated in trophoblast extracts (138b).

Several lines of experimental evidence thus indicate that the anti-inflammatory effects of steroid hormones may play a role in the immunologic coexistence of grávida and conceptus, especially considering the high placental concentrations of estrogen and progesterone. Even more intriguing is the conjecture that reproductive steroids may play a role in the shifting balance of thymic and lymphocyte proliferation and suppression as pregnancy evolves.

GESTATIONAL HORMONES Human chorionic gonadotropin (hCG) and human placental lactogen (hPL) are glycoproteins synthesized by the syncytiotrophoblast. Peak secretion rate (10^5 IU/day) and blood levels (10 IU/ml peripheral, 600 IU/gm placenta wet weight) of hCG are reached at ten-weeks gestation, but there is a progressive rise in hPL secretion throughout pregnancy. There is no evidence for an immunosuppressive effect of hPL. Initial studies with commercially available hCG extracted from pregnancy urine suggested an immunosuppressive role for hCG. In vivo, injection of crude hCG was shown to diminish delayed hypersensitivity to PPD in guinea pigs (138c), but did not interfere with rejection of intrauterine skin grafts in rats (138d). Crude lots of hCG in the dose range 1,000–10,000 IU/ml were noted to inhibit lymphocyte mitogenic response to PHA (138d, 138e, 139) or to allogeneic lymphocytes in MLR (140, 141, 142a). The suppressive effect was lost by washing lymphocytes preincubated with crude hCG (140) or by increasing the dose of PHA (142b). Several investigators

observed that low doses of crude or purified hCG (100–1000 IU/ml) enhanced the uptake of ^3H -thymidine by unstimulated lymphocytes in culture (142c, 143) or augmented lymphocyte responses to pokeweed mitogen (142c) or the MLR (142b, 143). However, hCG purified from the crude lots by Sephadex chromatography and gel filtration in higher doses of 10,000–15,000 IU/ml did not inhibit lymphocyte blastogenic responses to mitogens or allogeneic cells (142b, 143, 144).

Experiments have been carried out to define the immunosuppressive substances in the various preparations of crude hCG. Patillo et al (144) observed that concentrations of phenol equivalent to that in Organon hCG completely inhibited mitogenic responses to the plant lectins. Immunoabsorption of IgG removed the anticomplement activity from another preparation of crude hCG (145). Crude hCG can be separated by gel chromatography into fractions with biologic activity of hCG or inhibitory effects on lymphocyte transformation (146). Experiments with neuraminidase suggest that the inhibitory factors may be sialoglycoproteins found in pregnancy urine and distinct from hCG (143, 146).

ALPHA-FETOPROTEIN Alpha-fetoprotein (AFP) is a normal component of amniotic fluid and fetal serum in many species and has structural and biochemical similarities to albumin (147, 148). It is present in high concentrations early in fetal life and reaches a maximum concentration of several milligrams per milliliter in the first trimester of human pregnancy, falling gradually to 20–50 $\mu\text{g}/\text{ml}$ by birth (149). Murgita & Tomasi (150, 151) first reported that AFP isolated from mouse amniotic fluid was a potent inhibitor of in vitro mitogen-induced and mixed lymphocyte proliferation as well as primary and secondary humoral antibody responses. Since then a number of studies have presented conflicting findings (152–154). In these latter studies, lymphocyte proliferation to mitogens and antigens was either normal or abnormal, while the production of a lymphokine (macrophage migration inhibitory factor), E-rosette formation by T cells, and lymphocyte cytotoxicity were unaffected. Interestingly, AFP activates lymphocyte suppressor cells, and this finding may account in part for its inhibitory effects on cell function (155, 156). However, it is still unresolved whether this immunosuppressive property is due to AFP itself, a contaminant, or is an artifactual result of the process of purification, unrelated to its biologic effect in vivo (154).

PREGNANCY-SPECIFIC GLYCOPROTEINS In addition to hCG, there are a number of other pregnancy-associated glycoproteins that may have immunosuppressive properties. These include ovamucoid, fetuin, α_2 -macroglobulin and β_1 -glycoprotein (SP_1). α_2 -Macroglobulin isolated from maternal (not fetal) circulation has a molecular weight range of $3\text{--}7.5 \times 10^5$,

as reported by several laboratories (157–159), and is present on the surface of lymphocytes (160). It has inhibitory effects on MIF production and lymphoproliferative responses to mitogens and antigens (157, 161–163). Of interest, the inhibitory effects of α_2 -macroglobulin on mitogen responses have been attributed in part to its ability to bind to the lectin, thus limiting the quantity available to the cells (162). β_1 -Glycoprotein is a product of the syncytial trophoblast. Its concentration in serum increases with advancing pregnancy and may reach levels as high as 30 mg% (162). It has an apparent molecular weight of 1.1×10^5 and has been reported to suppress PHA responses (164). Further studies using more purified preparations are required to discern the exact role of these glycoproteins in suppressing maternal lymphocyte responses during pregnancy.

Specific Immunosuppression

SERUM While some studies show that maternal serum can nonspecifically inhibit cellular immune responses to several stimuli (165–167), many others demonstrate immunologic specificity. Such specific “enhancement” or “blocking” by maternal serum was first found in mice when the survival of paternal-strain tumors and skin grafts in females was increased by strain-specific pregnancies (168, 169). These effects became more profound with succeeding pregnancies but were usually detected with the first pregnancy. The latter findings prompted Kaliss & Dagg (170) to suggest that immunological enhancement, a phenomenon in which humoral antibodies facilitate the growth of antigenically foreign cells that are normally rejected in their absence, might be operative during pregnancy. This hypothesis was supported by the findings of Hellstrom et al (109), who showed that serum from BALB/C mice made pregnant by C3H males was able to prevent the in vitro inhibition of colony formation of C3H embryonic cells by presensitized BALB/C lymphocytes. Similar results were also obtained with tumors and could be passively transmitted (171).

In the in vitro systems, multiparous serum has been found to inhibit the MLR between maternal and paternal, maternal and newborn, and even between maternal and third-party cells with no absolute correlation with HLA type (105, 106, 172–174). Similar results have been obtained in other in vitro systems. As reported in the section on cellular immune responses, lymphocytes from multiparous women produce MIF in response to newborn or paternal cells and placental extracts (110, 111). This response can be specifically ablated by autologous maternal serum but not by serum from other multiparous women (175, 176). In addition, maternal lymphocytes can be prevented from exhibiting cytotoxicity toward trophoblast cells in vitro if the latter are grown in maternal serum (57).

The enhancing factor present in maternal serum that accounts for the effects described above has been partially characterized and studied for its possible biologic function in pregnancy. It has been described as belonging to the IgG class, since its elution pattern on gels is with 7S molecules and it can be removed by affinity chromatography with anti-IgG (57, 112, 174, 176). These antibodies can be absorbed by paternal cells but not by pooled human platelets, which suggests that they are not directed against the classical HLA-A, B, or C series antigens but against HLA-D antigens. Therefore, they may be analogous to the anti-Ia antibodies described in the mouse (102). These antibodies are cytotoxic to B cells (but not T cells), and their F(AB¹)₂ fragments suppress EA-rosette formation with Fc-bearing lymphocytes and inhibit the MLR (177).

A biologic role for these antibodies in the maintenance of a normal-term pregnancy has been suggested in clinical studies with women who have recurrent idiopathic spontaneous abortions. Rocklin et al (112) found that while lymphocytes from chronic aborters were presensitized to paternal antigens, their sera did not contain a blocking factor that could prevent their MIF response to paternal cells. In contrast, serum from normal multiparous women contained a blocking factor that did inhibit their MIF response to their mate. Of particular interest, blocking activity appeared in the serum of chronic aborters during subsequent successful pregnancies. In further studies, blocking activity has been detected in the serum of 80% of primagravid women between 16–24 weeks of pregnancy (131). Because blocking activity has not always been detected in the serum of normal primagravid women, one could question whether these antibodies play a significant role in fetal survival. One possible explanation for their absence from the serum of some normal women may be that they are attached to one possible target organ, the placenta, and are not present in high enough titer to be detected systemically by the available techniques. In fact, antibodies with blocking activity can be eluted from placentae, as is discussed below.

Although the data are suggestive, there is no objective proof that enhancing (blocking) antibodies play a major role in the *in vivo* survival of the mammalian embryo or fetus. However, in patients receiving allografts, a situation somewhat analogous to the fetal allograft, the presence of enhancing antibodies may improve graft survival. For example, hemodialysis patients who receive repeated blood transfusions develop blocking factors similar to those found in multiparous women (178). There is a better survival of future kidney grafts in these patients. Moreover, the administration of pooled IgG (obtained from retroplacental blood of multiparous women) to renal transplant patients has been associated with increased graft survival (compared to IgG obtained from normal serum) (179, 180).

IMMUNOGLOBULIN ELUTED FROM PLACENTAE Immunoglobulin has been identified on human trophoblast membrane and mouse placenta by immunohistologic techniques (61, 181). Placental material has been eluted and the IgG fraction isolated and tested for its immunosuppressive and enhancing activity. IgG eluted from human term placenta was shown to inhibit *in vitro* mitogen, antigen, and MLR proliferative responses (62, 182). This effect was nonspecific because the eluate inhibited normal as well as maternal cells. However, in more recent studies IgG eluted from human placenta was found to specifically suppress MIF production by maternal cells in response to paternal antigens (131). Of particular interest in this latter study, blocking activity was detected in the placental eluate but not in the serum of a primagravid woman, which supports the suggestion made earlier that blocking antibodies may be in low titer with first pregnancies. IgG eluted from mouse placenta binds *in vitro* to paternal lymphocytes and if administered *in vivo* will enhance the take of paternal-strain sarcoma allografts in mice of the maternal strain (181).

Analysis of the IgG eluted from pooled human placenta revealed at least two antibody populations similar to that present in multiparous sera (177). One antibody exhibits lymphocytotoxicity in the presence of complement and has anti-HLA specificities. The other has been identified as having specificity for HLA-D antigens and is comparable to the anti-Ia antibodies described earlier. It is of interest that alloantisera directed against mouse Ia and rat Ia-like antigens enhance skin or kidney allografts, and in man a similar antisera can impair *in vitro* antibody-dependent cellular cytotoxicity and Ig synthesis (102, 183, 184).

ROLE OF SUPPRESSOR CELLS In addition to their role as effector cells in cellular-immune reactions and as helper cells in antibody formation, T cells also play a regulatory role in the immune response (185). Both under *in vivo* and *in vitro* conditions, a subpopulation of T cells is capable of suppressing the function of killer cells, antibody-producing and lymphokine-producing cells. These cells can be identified in the mouse as having the Ly 2, 3⁺ phenotype (186) and in man as having Fc receptors for IgG (187) and histamine receptors (188a). Their role in pregnancy is being investigated (125).

HISTOCOMPATIBILITY AND PREGNANCY OUTCOME

The perhaps unwarranted model of the placenta as an allograft in the gravida naturally led to the proposition that antigenic differences between gravida and fetus might influence placental size. Billington first observed

that F₁ hybrids produced larger placentae than within-strain matings in the mouse (188b), and that there was greater proliferation of allogeneic than syngeneic trophoblast when transplanted to the testis (189). However, better controlled investigations in mice and other species produced no evidence that placental size or outgrowths were affected by major histocompatibility antigen differences (190–194). The initial report that immunization of mice prior to pregnancy increased placental weights (195) was also unconfirmed by subsequent studies (196). McLaren points out in her review of these studies (197) that larger hybrid placentae could result from the genetic process of heterosis, in which greater reproductive vigor parallels a higher degree of heterozygosity.

The question of the role of histoincompatibility and maternal immunologic status in determining fetal and placental weights remains open, however, because of the careful studies of Beer and co-workers (198, 199). They observed that mean placental weights and litter sizes were increased in allogeneic compared to syngeneic matings in hamsters, mice, and rats, and were further increased by presensitization of the gravidas. Allogeneic pregnancies also produced enlargement of maternal para-aortic lymph nodes and spleens as another indicator of immune response. Excision of para-aortic nodes prior to mating in allogeneic pregnancies resulted in smaller fetal-placental units.

A few studies found decreased litter size from embryonic mortalities subsequent to maternal preimmunization to paternal antigens (196, 200). This result might be explained by the reduction in the maternal decidual reaction produced in mice by histoincompatible matings (201) or maternal immunization (202). The reduced decidual reaction was associated with smaller fetuses, but not placentae, in those studies.

There is little evidence of immunologic influences on fetal growth or placental size in human pregnancy, although large-scale systematic studies are yet to be performed. ABO incompatibility has not been confirmed as a determinant of placental growth (197). Maternal production of anti-HLA antibodies is not associated with adverse pregnancy outcome, except for an unconfirmed association of fetal congenital anomalies with maternal leukocytotoxic antibodies (203). Heightened maternal-fetal MLR was supposedly associated with larger placentae in one preliminary study (172). Since typing of cells and sera for minor histocompatibility loci is now possible, our knowledge of the immunogenetics of human reproduction should advance in the next few years.

Preeclampsia is a syndrome of acute hypertension and glomerular damage primarily arising in the latter half of first pregnancies. Since the cause(s) remains unknown, many have sought evidence (so far unsuccessfully) that some pathologic features of the syndrome are immunologically mediated.

These studies were recently reviewed (204). A greater number of histocompatible pairings among women with preeclampsia was found (205), but not confirmed elsewhere (206). In a preliminary study, Bodmer found that a surprising percentage of subjects with severe preeclampsia were homozygous at both HLA loci (207). The implications of this remain unknown.

FETAL IMMUNOLOGIC RESPONSES TO MATERNAL ANTIGEN

Despite the report in 1904 (208) of the ability of a fetal goat to produce antibodies to heterologous blood injected in utero, the fetus was considered to be immunologically incompetent until, in the early 1950s, Australian workers demonstrated that a fetal lamb could reject a skin allograft in utero (209). Recent work in the mouse, lamb, and primate has done much to quantify what the nineteenth century pathologists described in congenital infections of the newborn.

Competence of immune responses arises early in development, despite the small numbers of immunocompetent cells; this raises the possibility of multipotentiality of the cells. Immunocompetence develops in a controlled manner with humoral responses to certain antigens developing in a stepwise fashion. The maturation of immunologic competence occurs prior to maturation of lymph nodes, spleen, or gut-associated nodes. This presumably arises in the lymphoid cells of the fetal liver, the mammalian equivalent of the bursa of Fabricius of birds (210). The initial fetal immune responses are mature in quantity and quality of cellular responses and immunoglobulin titers with class sequences IgM, IgG₁ and IgG₂ (211).

A recent report suggests that one function of the thymus in the developing embryo may be its ability to alter bone-marrow-derived lymphocytes in such a way that on leaving the thymus, the lymphocytes will recognize their own "self" antigens (212). Interestingly, the thymus of mouse embryos will reconstitute the immune responses of their thymectomized mother following pregnancy (213).

The inability of neonates to respond to some antigens when challenged early appears to be the result of suppression by passively acquired maternal IgG rather than a defect in the neonatal immune system (214, 215).

SUMMARY

It is clear from the foregoing that no one simple mechanism can possibly explain the success of the fetus as an allograft. In fact, it would be surprising if there were just one, considering the importance of reproduction for the propagation and survival of the species. There appear to be an array of

mechanisms that probably contribute to this process. Some may be specific in nature, while others are nonspecific.

The maternal immune system is challenged with paternal antigens through exposure to trophoblast tissue and fetal cells crossing the placenta into the maternal circulation. The dose of antigen, the manner of presentation (cellular, subcellular, or soluble), and the nature of the antigen all determine the type of response that will be elicited. It is also clear that complex maternal immunologic responses, including antibodies to red cell antigens, HLA-A, B, and C series and HLA-D antigens, and cell-mediated responses such as proliferation, lymphokines, cytotoxicity, and possibly suppressor cells, are generated to a variety of paternal antigenic determinants. The fact that some of these reactions are detected *in vitro* in the absence of maternal serum, but not in its presence, suggests that the local milieu is important in influencing their expression *in vivo*. For example, such factors as hormones (cortisol, progesterone, estrogen), pregnancy-associated glycoproteins (α_2 -macroglobulin, β_1 -glycoprotein), and AFP, which have immunosuppressive properties, may all serve to nonspecifically inhibit and decrease the general tone of maternal immunologic responses, particularly at the placental interface where many of these are present in high concentrations. However, these nonspecific factors may not be sufficient to prevent presensitized effector lymphocytes from continuing an ongoing rejection process, as is often the case in the chronic rejection of an allograft. For this purpose specific enhancing antibodies would play an important role.

There may be a subtle balance created on the trophoblast cell surface between specific antibodies and trophoblast or embryonic alloantigens resulting in a limited expression of antigens capable of inducing rejection reactions. This could favor the production of blocking antibodies and/or T suppressor cells as opposed to cytotoxic antibody and killer cells. In fact, low levels of antigen density on cell surface favor a blocking effect by IgG rather than cytotoxicity (216).

Blocking or enhancing antibodies might exert their effect on maternal immunologic responses in several ways. They could block the afferent limb by combining with antigen and preventing sensitization or increasing the level of sensitivity. An example of the latter would be the coating of fetal cells that enter the maternal circulation. Enhancing antibodies could work directly on the effector cells to suppress their function. The antibody itself, or more likely, antigen-antibody complexes, may be important in this regard. It is known that antigen-antibody complexes can suppress effector lymphocytes in tumor-bearing individuals (217), and have been detected in the circulation of pregnant mice and guinea pigs (218). The effector limb of the maternal response could also be regulated by enhancing antibodies.

Adsorption of the antibodies to trophoblast tissue could render it less susceptible to immunologic attack. Furthermore, paternal antigens could be "covered" or shed into the circulation, thus preventing the recognition of the tissue as foreign. Therefore, despite the presence of sensitized maternal cells, no reaction would ensue because they would not be triggered.

While the role of blocking antibodies is strongly suggested as a mechanism of fetal survival, formal proof is still lacking. There is an association between the presence of these antibodies in normal multigravid women and their absence in women who chronically abort. However, their presence in primigravid women is not consistent. Therefore, if they do play an important role one must explain a successful first pregnancy in their absence. As suggested previously, it is possible that the antibodies are fixed in the placenta but are not present in high enough titer to be detected systemically. In fact, IgG with blocking activity has been eluted from placentae. Under other circumstances the administration of these anti-Ia-like antibodies in vivo has also been shown to enhance allograft survival. Taken together, the data strongly implicate these antibodies as one mechanism of survival of the fetus. It would be of considerable interest to determine whether IgG obtained from normal multigravid women would provide protection to the fetus of women who chronically abort. However, further characterization of these antibodies must occur before such studies can be undertaken.

Literature Cited

1. Medawar, P. B. 1954. Some immunological and endocrinological problems raised by the evolution of viviparity in vertebrates. *Symp. Soc. Exp. Biol.* 7: 320-28
2. Jones, W. R. 1976. Immunological aspects of infertility. In *Immunology of Human Reproduction*, ed. J. S. Scott, W. R. Jones, pp. 375-413. London: Academic
3. Fellous, M., Daussett, J. 1970. Probable haploid expression of HLA antigens on human spermatozoa. *Nature* 225: 191-93
4. Erickson, R. P. 1977. Differentiation and other alloantigens of spermatozoa. In *Immunology of Gametes*, ed. M. Edidin, M. H. Johnson, pp. 85-107. London: Cambridge Univ. Press
5. Goldberg, E. 1974. Effects of immunization with LHH-X on fertility. In *Immunology Approaches to Fertility Control*, ed. E. Diezfallusz, pp. 202-22. Stockholm: Karolinska Institute
6. Cohen, J., Werrett, D. J. 1975. Antibodies and sperm survival in the female tract of the mouse and rabbit. *J. Reprod. Fertil.* 42:301-10
7. Cohen, J., McNaughton, D. C. 1974. Spermatozoa: the probable selection of a small population by the genital tract of the female rabbit. *J. Reprod. Fertil.* 39:297-310
8. Overstreet, J. W., Katz, D. F. 1977. In *Development in Mammals*, ed. M. H. Johnson, pp. 31-66. Amsterdam: North-Holland
9. Beer, A. E., Billingham, R. E. 1974. Host responses to intra-uterine tissue, cellular and fetal allografts. *J. Reprod. Fertil.* 21:59-88
10. Vacquier, V. D. 1977. Hooks and eyes of sperm and eggs. *Sci. News* 112:356
11. Yanagimachi, R. 1977. Specificity of sperm-egg interaction. See Ref. 4, pp. 255-89
12. Kirby, D. R. S. 1969. On the immunologic function of the zona pellucida. *Fertil. Steril.* 20:933-37
13. Bernard, O., Ripoché, M. A., Bennett, D. 1977. Distribution of maternal immunoglobulins in the mouse uterus and embryo in the days after implantation. *J. Exp. Med.* 145:58-75
14. Sellens, M. H., Jenkinson, E. J. 1975. Permeability of the mouse zona pel-

- lucida to immunoglobulin. *J. Reprod. Fertil.* 42:153-57
15. Kirby, D. R. S. 1968. The immunological consequences of extra-uterine development of allogeneic mouse blastocysts. *Transplantation* 6:1005-9
16. Searle, R. F., Johnson, M. H., Billington, W. D. 1974. Investigation of H-2 and non-H-2 antigens on the mouse blastocyst. *Transplantation* 18:136-41
17. Billington, W. D. 1976. The immunobiology of trophoblast. See Ref. 2, pp. 81-102
18. Fauve, R. M., Hevim, B., Jacob, H. 1974. Anti-inflammatory effects of murine malignant cells. *Proc. Natl. Acad. Sci. USA* 71:4052-56
19. Heyner, S. 1973. Detection of H-2 antigens on cells of the early mouse embryo. *Transplantation* 16:675-78
20. Carter, J. 1976. The effect of progesterone, estradiol, and hCG on cell-mediated immunity in pregnant mice. *J. Reprod. Fertil.* 46:211-16
21. Billington, W. D., Jenkinson, J., Searle, R. F., Sellens, M. H. 1977. Alloantigen expression during early embryogenesis and placental ontogeny in the mouse: immunoperoxidase and mixed hemoadsorption studies. *Transplant. Proc.* 9:1371-77
22. Searle, R. F., Sellens, M. H., Elson, J., Jenkinson, E. J., Billington, W. D. 1976. Detection of alloantigens during pre-implantation development and early trophoblast differentiation in the mouse by immunoperoxidase labelling. *J. Exp. Med.* 143:348-59
23. Webb, C. G., Gall, W. E., Edelman, G. M. 1977. Synthesis and distribution of H-2 antigens in pre-implantation mouse embryos. *J. Exp. Med.* 146:923-32
24. Hakansson, S., Heyner, S., Sundqvist, K. G., Bergstrom, S. 1975. The presence of paternal H-2 antigens on hybrid mouse blastocysts during experimental delay of implantation and the disappearance of their antigens after onset of implantation. *Int. J. Fertil.* 20:137-40
25. Ledford, B. E., Rankin, J. E., Froble, L., Serve, M. J., Markwald, R. R., Baggett, B. 1978. The decidual cell reaction in the mouse uterus: DNA synthesis and autoradiographic analysis of responsive cells. *Biol. Reprod.* 18:506-9
26. Ansell, J. D., Barlow, P. W., McLaren, A. 1974. Binucleate and polyploid cells in the decidua of the mouse. *J. Embryol. Exp. Morphol.* 31:223-27
27. Gardner, R. L. 1975. Origins and properties of trophoblast. In *Immunobiology of Trophoblast*, ed. R. G. Edwards, C. W. S. Howe, M. H. Johnson, pp. 43-61. Cambridge Univ. Press
28. Chapman, V. M., Ansell, J. D., McLaren, A. 1972. Trophoblast giant cell differentiation in the mouse: expression of glucose phosphate isomerase (GP1-1) electrophoretic variants in transferred and chimeric embryos. *Dev. Biol.* 29:48-54
29. Hill, J. P. 1900. On the fetal membranes, placenta and parturition of the Native cat. *Anat. Anz.* 18:364-73
30. Hill, J. P. 1897. The placenta of Perameles (Contributions to the embryology of the Marsupialia I). *Q. J. Microsc. Sci.* 40:385-446
31. Hill, J. P. 1900. Contributions to the embryology of the Marsupialia II. On a further stage in the placenta of Perameles. *Q. J. Microsc. Sci.* 43:1-23
32. Park, W. W. 1968. Fetal maternal cellular interactions. *J. Reprod. Fertil.* 3:37-39
33. Enders, A. C. 1976. Cytology of human early implantation. *Res. Reprod.* 8:1
34. Padykula, H. A., Taylor, J. M. 1977. In *Reproduction and Evolution*, Proc. Int. Symp. Comp. Biol. Reprod., 4th, eds. J. H. Calaby, C. H. Tyndale-Biscoe, pp. 303-24. Australian Acad. Sci.
35. Larsen, J. F. 1961. Electron microscopy of the implantation site in the rabbit. *Am. J. Anat.* 109:319-25
36. Enders, A. C., Schlafke, S. 1971. Penetration of the uterine epithelium during implantation in the rabbit. *Am. J. Anat.* 132:219-40
37. McLennan, J. G. 1974. Ultrastructural studies of early nidation in pregnancy and pseudopregnancy. *Am. J. Obstet. Gynecol.* 120:319-34
38. Tekelioglu-Uysal, M., Edwards, R. G., Kisinisci, H. A. 1975. Ultrastructural relationships between decidua, trophoblast and lymphocytes at the beginning of human pregnancy. *J. Reprod. Fertil.* 42:431-38
39. Galassi, L. 1967. Reutilization of maternal nuclear material by embryonic trophoblastic cells in the rat for the synthesis of DNA. *J. Histochem. Cytochem.* 15:573-79
40. Ledoux, L., Charles, P. 1967. Uptake of exogenous DNA by mouse embryos. *Exp. Cell Res.* 45:498-501
41. Kaye, M. D. 1977. An autoradiographic study of the implantation of transferred mouse blastocysts. *Aust. J. Biol. Sci.* 30:577-82
42. Harris, H., Sidebottom, E., Grace, D. M., Bramwell, M. E. 1969. The expression of genetic information, a study

- with hybrid animal cells. *J. Cell Sci.* 4:499-525
43. Croce, C. M., Kaprowski, H. 1978. The genetics of human cancer. *Sci. Am.* 238:117
 44. Gearhart, J. D., Mintz, B. 1972. Glucose phosphate isomerase subunit-reassociation tests for maternal-fetal and fetal-fetal cell fusion in the mouse placenta. *Dev. Biol.* 29:55
 45. Koren, Z. 1968. Antigenicity of mouse placental tissue. *Am. J. Obstet. Gynecol.* 102:340-46
 46. Beer, A. E., Billingham, R. E., Yang, S. L. 1972. Further evidence concerning the autoantigenic status of the trophoblast. *J. Exp. Med.* 135:1177-84
 47. Youtananukorn, V., Matangkasombut, P., Osathanondh, V. 1974. Onset of human maternal cell-mediated immune reaction to placental antigens during the first pregnancy. *Clin. Exp. Immunol.* 16:593-98
 48. Goodfellow, P. N., Barnstable, C. J., Bodmer, W. F., Snary, D., Crumpton, M. J. 1976. Expression of HLA system antigens on placenta. *Transplantation* 22:595-603
 49. Kirby, D. R. S., Billington, W. D., James, D. A. 1966. Transplantation of eggs to the kidney and uterus of immunized mice. *Transplantation* 4:713-18
 50. Simmons, R. L., Russell, P. S. 1962. The antigenicity of mouse trophoblast. *Ann. NY Acad. Sci.* 99:717-32
 51. Beer, A. E., Head, J. R., Smith, W. G., Billingham, R. E. 1976. Some immunoregulatory aspects of pregnancy in rats. *Transplant. Proc.* 8:267-73
 52. Bradbury, S., Billington, W. D., Kirby, D. R. S., Williams, F. A. 1969. Surface mucin of human trophoblast. *Am. J. Obstet. Gynecol.* 104:416-18
 53. Currie, G. A., Van Doorninck, W., Bagshawe, K. O. 1968. Effect of neuraminidase on the immunogenicity of early mouse trophoblast. *Nature* 219:191
 54. Simmons, R. L., Lipschultz, M. L., Rios, A., Ray, P. K. 1971. Failure of neuraminidase to unmask histocompatibility antigens on trophoblast. *Nature* 231:111-12
 55. Loke, Y. W., Josey, V. C., Borland, R. 1971. HLA antigens on human trophoblast cells. *Nature* 232:403
 56. Douthwaite, R. M., Uhrbach, G. I. 1971. In vitro antigenicity of trophoblast. *Am. J. Obstet. Gynecol.* 109:1023-28
 57. Taylor, P. V., Hancock, K. W. 1975. Antigenicity of trophoblast and possible antigen-masking effect during pregnancy. *Immunology* 28:973-82
 58. Jenkinson, E. J., Billington, W. D. 1974. Differential susceptibility of mouse trophoblast and embryonic tissue to immune cell lysis. *Transplantation* 18:286-88
 59. Johnson, P. M., Faulk, W. P. 1978. Immunological studies of human placentae: identification and distribution of proteins in immature chorionic villi. *Immunology* 34:1027-35
 60. Fox, H. 1968. Fibrinoid necrosis of placental villi. *J. Obstet. Gynaecol. Br. Commonw.* 75:448-52
 61. McCormick, J. N., Faulk, W. P., Fox, H., Fudenberg, H. H. 1971. Immunohistological and elution studies of the human placenta. *J. Exp. Med.* 133:1-12
 62. Faulk, W. P., Jeannot, M., Creighton, W. D., Carbonara, A., Hay, F. 1974. Immunological studies of the human placenta. *J. Clin. Invest.* 54:1011-19
 63. Faulk, W. P., Johnson, P. N. 1977. Immunological studies of human placentae: identification and distribution of proteins in mature chorionic villi. *Clin. Exp. Immunol.* 27:365-75
 64. Faulk, W. P., Sanderson, A., Temple, A. 1977. Distribution of MHC antigens in human placental chorionic villi. *Transplant. Proc.* 9:1379-84
 65. Billingham, R. E. 1978. From transplantation biology to reproductive immunobiology. *Obstet. Gynecol. Ann.* 7:1-14
 66. Park, W. V. 1971. In *Choriocarcinoma: A Study of Its Pathology*. London: Heineman
 67. Edwards, R. G. 1975. Discussion. See Ref. 27, p. 149
 68. Wynn, R. M. 1976. Foetomaternal cellular relations in the human basal plate. An ultrastructural study of the placenta. *Am. J. Obstet. Gynecol.* 97:832
 69. Nelson, J. H., Hall, J. E., Manuel-Limson, G. 1967. Effect of pregnancy on the thymolymphatic system. *Am. J. Obstet. Gynecol.* 98:895
 70. Nelson, J. H., Lo, T., Hall, J. E., Krown, S., Nelson, J. M., Fox, C. W. 1973. The effect of trophoblast on immune state of women. *Am. J. Obstet. Gynecol.* 117:689-99
 71. Head, J. R., Hamilton, M. S., Beer, A. E. 1978. Maternal hamster immune response to alloantigens of fetus. *Fed. Proc.* 37:2054-56
 72. Adinolfi, M. 1975. The human placenta as a filter for cells and plasma proteins. See Ref. 27, pp. 193-210

73. Schroder, J., De La Chapelle, A. 1972. Fetal lymphocytes in the maternal blood. *Blood* 39:153
74. Schroder, J. 1975. Are fetal cells in maternal blood mainly B lymphocytes? *Scand. J. Immunol.* 4:279-85
75. Walknowska, J., Conte, F. A., Grumbach, M. M. 1969. Practical and theoretical implications of fetal/maternal lymphocyte transfer. *Lancet* 1:1119-22
76. Brambell, F. W. R. 1966. The transmission of immunity from mother to young and the catabolism of immunoglobulins. *Lancet* 2:1087-93
77. Lanman, J. T., Herod, L. 1965. Homograft immunity in pregnancy. The placental transfer of cytotoxic antibody in rabbits. *J. Exp. Med.* 122:579-86
78. Carlson, G. A., Wegmann, T. G. 1978. Paternal-strain antigen excess in semi-allogeneic pregnancy. *Transplant. Proc.* 10:403-7
79. Morisada, M., Yamaguchi, H., Iizuka, R. 1976. Immunobiological function of syncytiotrophoblast: a new theory. *Am. J. Obstet. Gynecol.* 125:3-16
80. Johnson, P. M., Trenchev, P., Faulk, P. W. 1975. Immunological studies of human placenta binding of complexed immunoglobulin by stromal endothelial cells. *Clin. Exp. Immunol.* 22:133-38
81. Jenkinson, E. J., Billington, W. D. 1977. Cell surface properties of early mammalian embryos. In *Concepts in Mammalian Embryogenesis*, ed. M. I. Sherman, pp. 235-66. Cambridge: MIT Press
82. Edeidin, M. 1972. Histocompatibility genes in transplantation antigens and pregnancy. In *Transplantation Antigens: Markers of Biological Individuality*, ed. B. D. Kahan, R. A. Rensfeld, pp. 75-114. New York: Academic
83. Queenan, J. T. 1977. *Modern Management of the RH Problem*, p. 9. Hagerstown: Harper & Row. 2nd ed.
84. van Rood, J. J., van Leeuwen, A., Eernisse, J. G. 1959. Leukocyte antibodies in sera of pregnant women. *Vox Sang.* 4:427-44
85. Payne, R. 1962. The development and persistence of leukoglutinin in parous women. *Blood* 19:411-24
86. Lawler, S. D., Klonda, P. T., Bagshawe, K. D. 1974. Immunogenicity of molar pregnancies in the HLA system. *Am. J. Obstet. Gynecol.* 120:857-61
87. Ahrons, S. 1971. Leukocyte antibodies: occurrence in primigravidae. *Tissue Antigens* 1:178-83
88. Finn, R. S., Hill, C. A., Gover, A. J., Rolfs, I. G., Gurney, I. J., Denye, V. 1972. Immunological responses in pregnancy and survival of fetal homograft. *Br. Med. J.* 111:150-53
89. Strelkauskas, A. J., Wilson, B. S., Dray, S., Dodsden, M. 1975. Inversion of human T and B cell levels in early pregnancy. *Nature* 258:331-33
90. Gudson, J. P. Jr. 1969. Fetal and maternal immunoglobulin levels during pregnancy. *Am. J. Obstet. Gynecol.* 103: 895-900
91. Maroulis, G. B., Buckley, R. H., Younger, J. B. 1971. Serum immunoglobulin concentrations during normal pregnancy. *Am. J. Obstet. Gynecol.* 109:971-76
92. Gudson, J. P. Jr., Prichard, D. 1972. Immunoglobulin D in pregnancy. *Am. J. Obstet. Gynecol.* 112:867-73
93. Valquist, B., Lagecrantz, R., Nordbring, F. 1950. Maternal and fetal titres of antistreptolysin and antistaphylolysin at different stages of gestation. *Lancet* 2:851-53
94. Lichtenstein, M. R. 1942. Tuberculin reaction in tuberculosis during pregnancy. *Am. Rev. Tuberc.* 49:89-92
95. Montgomery, W. P., Young, R. C., Aller, M. D., Horden, K. A. 1968. The tuberculin test in pregnancy. *Am. J. Obstet. Gynecol.* 100:829-31
96. Peer, L. A. 1958. Behavior of skin grafts exchanged between parents and offspring. *Ann. NY Acad. Sci.* 73:854-59
97. Kitzmiller, J. L., Stoneburger, L., Yelenosky, P. F., Lucas, W. E. 1973. Serum complement in normal pregnancy and pre-eclampsia. *Am. J. Obstet. Gynecol.* 117:312-15
98. Mitchell, G. W., Jacobs, A. A., Hadad, V., Paul, B. B., Straus, R. R., Sbama, A. J. 1970. The role of the phagocyte in host-parasite interactions. XXV. Metabolic and bacterial activities of leukocytes from pregnant women. *Am. J. Obstet. Gynecol.* 108:805-13
99. Van der Werf, A. J. M. 1971. Are lymphocytotoxic iso-antibodies induced by early human trophoblast? *Lancet* 1:595-96
100. Doughty, R. W., Gelsthorpe, K. 1974. An initial investigation of lymphocyte antibody activity through pregnancy and in eluates prepared from placental material. *Tissue Antigens* 4:291-98
101. Winchester, R. J., Fu, S. M., Wernet, P., Kunke, H. G., Dupont, B., Jersild, C. 1975. Recognition by pregnancy serums of non-HLA alloantigens selectively expressed on B lymphocytes. *J. Exp. Med.* 141:924-29

102. Davies, D. A. L., Staines, N. A. 1976. A cardinal role for I-region antigens (Ia) in immunological enhancement, and the clinical implications. *Transplant. Rev.* 30:18-39
103. Lewis, J. L., Slang, L., Nagel, B., Oppenheim, J. J., Perry, S. 1966. Lymphocyte transformation in mixed leukocyte cultures in women with normal pregnancy or tumors of placental origin. *Am. J. Obstet. Gynecol.* 96: 287-90
104. Cepellini, R., Bonnard, G. D., Coppo, F., Migliano, V. D., Popisil, M., Curtoni, E. S., Pelligrino, M. 1971. Mixed leukocyte cultures and HLA antigens. I. Reactivity of young fetuses, newborns and mothers at delivery. *Transplant. Proc.* 3:58-63
105. Revillard, J. P., Robert, M., Betuel, H., Latour, M., Bonneau, M., Brochier, J., Trager, J. 1972. Inhibition of the mixed lymphocyte reaction by antibodies. *Transplant. Proc.* 4:173-76
106. Robert, M., Betuel, H., Revillard, J. P. 1973. Inhibition of the mixed lymphocyte reaction by sera from multipara. *Tissue Antigens* 3:39-56
107. Olding, L. B., Oldstone, M. B. A. 1974. Lymphocytes from human newborns abrogate mitosis of their mother's lymphocytes. *Nature* 249:161-63
108. Wolf, R. L., Lomnitzer, R., Rabson, A. R. 1977. An inhibition of lymphocyte proliferation and lymphokine production released by unstimulated fetal monocytes. *Cell. Immunol.* 27:464-68
109. Hellstrom, K. E., Hellstrom, I., Brown, J. 1969. Abrogation of cellular immunity to antigenically foreign mouse embryonic cells by a serum factor. *Nature* 224:914-15
110. Youtananukorn, V., Matangkasombu, P. 1972. Human maternal cell mediated immune reaction to placental antigens. *Clin. Exp. Immunol.* 11:549-56
111. Rocklin, R. E., Zuckerman, J. E., Alpert, E., David, J. R. 1973. Effect of multiparity on human maternal hypersensitivity to fetal antigen. *Nature* 241:130-32
112. Rocklin, R. E., Kitzmiller, J. L., Carpenter, C. B., Garovoy, M. R., David, J. R. 1976. Maternal-fetal relation: absence of an immunologic blocking factor from the serum of women with chronic abortions. *N. Engl. J. Med.* 295:1209-13
113. Tulchinsky, D. 1973. Placental secretion of unconjugated estrone, estradiol, and estriol into the maternal and the fetal circulation. *J. Clin. Endocrinol. Metab.* 36:1079-87
114. Krohn, P. L. 1954. The effect of steroid hormones on the survival of skin homografts in the rabbit. *J. Endocrinol.* 11:78-82
115. Hulka, J. F., Mohr, K. 1967. Interference of cortisone-induced homograft survival by progestins. *Am. J. Obstet. Gynecol.* 97:407-10
116. Simmons, R. L., Price, A. L., Ozerkis, A. J. 1968. The immunologic problems of pregnancy. V. The effect of estrogen and progesterone on allograft survival. *Am. J. Obstet. Gynecol.* 100:908-11
117. Waltman, S. R., Burde, R. M., Barrios, J. 1971. Prevention of corneal homograft rejection by estrogens. *Transplantation* 11:194-96
118. Thompson, J. S., Crawford, M. K., Reilly, R. W., Severson, C. D. 1967. The effect of estrogenic hormones on immune responses in normal and irradiated mice. *J. Immunol.* 98:331-35
119. Weinstein, Y., Lindner, H. R., Eckstein, B. 1977. Thymus metabolizes progesterone—possibly enzymatic marker for T lymphocytes. *Nature* 266:632-33
- 120a. Tulchinsky, D., Okada, D. M. 1975. Hormones in human pregnancy. IV. Plasma progesterone. *Am. J. Obstet. Gynecol.* 121:293-99
- 120b. Ferre, F., Janssens, Y., Tanguy, G., Breuiller, M., DePariente, D., Cedard, L. 1978. Steroid concentrations in human myometrial and placental tissues at week 39 of pregnancy. *Am. J. Obstet. Gynecol.* 131:500-2
121. Hulka, J. F., Mohr, K., Liberman, M. W. 1965. Effect of synthetic progestational agents on allograft rejections and circulating antibody production. *Endocrinology* 77:897-901
122. Munroe, J. S. 1971. Progesteroids as immunosuppressive agents. *J. Reticuloendothel. Soc.* 9:361-75
123. Pettirossi, H. A., Wechter, W. J., Kountz, S. L. 1976. Prolongation of rat heart allografts survival by a synthetic progestagen (melengestrol acetate and ara-cytidine acylates). *Transplantation* 21:408-11
124. Siiteri, P. K., Febres, F., Clemens, L. E. 1977. Progesterone in maintenance of pregnancy: is progesterone nature's immunosuppressant? *Ann. NY Acad. Sci.* 286:384-97
125. Beer, A. E., Billingham, R. E. 1979. Maternal immunological recognition mechanisms during pregnancy. In *Maternal Recognition Mechanisms*. CIBA Found. Symp., Vol. 64. In press

126. Ablin, R. M., Bruns, G. R., Guinan, P. 1974. The effect of estrogen on the incorporation of ^3H -thymidine by PHA-stimulated human peripheral blood. *J. Immunol.* 113:705-7
127. Mori, T., Kobayashi, H., Nishimoto, H. 1977. Inhibitory effect of progesterone and 20α -hydroxypregn-4-en-3-one on the phytohemagglutinin-induced transformation of human lymphocytes. *Am. J. Obstet. Gynecol.* 127:151-57
128. Wyle, F. A., Kent, J. R. 1977. Immunosuppression by sex steroid hormones. *Clin. Exp. Immunol.* 27:407-15
129. Schiff, R. I., Mercier, D., Buckley, R. H. 1975. Inability of gestational hormones to account for the inhibitory effects of pregnancy plasma on lymphocyte responses in vitro. *Cell. Immunol.* 20:69-80
130. Clemens, L. E., Stites, D., Siiteri, P. K. 1977. In vitro suppression of human lymphocyte transformation by progesterone. *Gynecol. Invest.* 8:20
131. Kitzmiller, J. L., Rocklin, R. E. 1978. Unpublished data
132. Kenny, J. F., Pangburn, P. O., Trail, G. 1976. Effect of estradiol on immune competence: in vivo and in vitro studies. *Infect. Immun.* 13:448-56
133. Peterson, R. E., Imperato-McGinley, J. 1976. Cortisol metabolism in the perinatal period. In *Diabetes and Other Endocrine Disorders During Pregnancy and in the Newborn*, ed. M. I. New, R. H. Fiser, Jr., pp. 141-172. New York: Liss
134. Claman, H. N. 1972. Corticosteroids and lymphoid cells. *N. Engl. J. Med.* 287:388-97
135. Schreiber, A. D. 1977. Clinical immunology of the corticosteroids. *Prog. Clin. Immunol.* 3:103-14
136. Makman, M. H., Dvorkin, B., White, A. 1968. Influence of cortisol on the utilization of precursors of nucleic acids and protein by lymphoid cells in vitro. *J. Biol. Chem.* 243:1485-97
137. Balow, J. E., Rosenthal, A. S. 1973. Glucocorticoid suppression of macrophage migration inhibitory factor. *J. Exp. Med.* 137:1031-41
- 138a. Weston, W. L., Claman, H. N., Krueger, G. G. 1973. Site of action of cortisol in cellular immunity. *J. Immunol.* 110:880-83
- 138b. Werthamer, S. F., Govindaraj, S., Amaral, L. 1976. Placenta, transcortin, and localized immune response. *J. Clin. Invest.* 57:1000-8
- 138c. Han, T. 1975. Human chorionic gonadotropin. Its inhibitory effect on cell mediated immunity in vivo and in vitro. *Immunology* 29:509-15
- 138d. Kaye, M. D., Jones, W. R., Ing, R. M. Y., Markham, R. 1971. Effect of human chorionic gonadotropin on intrauterine skin allograft survival in rats. *Am. J. Obstet. Gynecol.* 110:640-43
- 138e. Kaye, M. D., Jones, W. R. 1971. Effect of hCG in vitro lymphocyte transformation. *Am. J. Obstet. Gynecol.* 109:1029-31
139. Contractor, S. F., Davies, H. 1973. Effect of human chorionic somatomammotrophin and hCG on phytohemagglutinin-induced lymphocyte transformation. *Nature* 243:284
140. Adcock, E. W. III, Teasdale, E., August, C. S., Cox, S., Meschia, G., Battaglia, F. C., Naughton, M. A. 1973. Human chorionic gonadotrophin: its possible role in maternal lymphocyte suppression. *Science* 181:845-47
141. Han, T. 1974. Inhibitory effect of human chorionic gonadotrophin on lymphocyte blastogenic responses: hCG and mitogen, antigen and allogenic human cells. *Clin. Exp. Immunol.* 18:529-35
- 142a. Jenkins, D. M., Acres, M. G., Peters, J., Riley, J. 1972. Human chorionic gonadotropin and the fetal allograft. *Am. J. Obstet. Gynecol.* 114:13-15
- 142b. Caldwell, J. L., Stites, D. P., Fudenberg, H. H. 1975. Human chorionic gonadotropin: effects of crude and purified preparations on lymphocyte responses to phytohemagglutinin and allogeneic stimulation. *J. Immunol.* 115:1249
- 142c. Beck, D., Ginsburg, H., Naot, Y. 1977. Modulating effect of human chorionic gonadotropin on lymphocyte blastogenesis. *Am. J. Obstet. Gynecol.* 129:14-20
143. Morse, J. H., Stearns, G., Arden, J. 1976. The effects of crude and purified human gonadotropin on in vitro stimulated human lymphocyte cultures. *Cell. Immunol.* 25:178-88
144. Pattillo, R. A., Shalaby, M. R., Hussa, R. O., Bahl, O. M. P., Mattingly, R. F. 1976. Effect of crude and purified hCG on lymphocyte blastogenesis. *Obstet. Gynecol.* 47:557-61
145. Loke, Y. W., Pepys, M. B. 1975. Effects of human chorionic gonadotropin preparations on complement in vitro. *Am. J. Obstet. Gynecol.* 121:37
146. Merz, W. E., Schmidt, W., Hilgenfeldt, U., Schackert, K., Lenhard, V. 1976. Isolation of substances from crude preparation of human chorionic gonadotro-

- pin (hCG) which strongly inhibit the transformation of lymphocytes. *Z. Immunitaetsforsch.* 152:286-89
147. Alpert, E., Drysdale, J. W., Schur, P. H., Isselbacher, K. J. 1971. Human AFP: purification and physical properties. *Fed. Proc.* 30:246
 148. Ruoslahti, E., Terry, W. D. 1976. Alpha-fetoprotein and serum albumin show sequence homology. *Nature* 260:804-5
 149. Gitlin, D., Biasucci, A. 1969. Development of IgG, IgA, IgM, β IC, β 1a, C1, esterase inhibition, ceruloplasmin, transferrin, haptoglobin, fibrinogen, plasminogen, orosomucoid, α 2-macroglobulin and pre-albumin in the conceptus. *J. Clin. Invest.* 48:1433-66
 150. Murgita, R. A., Tomasi, T. B. 1975. Suppression of the immune response by alpha-fetoprotein. I. Primary and secondary response. *J. Exp. Med.* 141: 269-86
 151. Murgita, R. A., Tomasi, T. B. 1975. Suppression of the immune response by alpha-fetoprotein. II. MLR and mitogen-induced lymphocyte transformation. *J. Exp. Med.* 141:440-52
 152. Sheppard, H. W., Sell, S., Trefts, P., Bahu, R. 1977. Effects of alpha-fetoprotein on murine immune responses. I. Studies on mice. *J. Immunol.* 119: 91-97
 153. Sell, S., Sheppard, H. W., Poler, M. 1977. Effects of alpha-fetoprotein on murine immune response. II. Studies on rats. *J. Immunol.* 119:98-103
 154. Littman, B. H., Alpert, E., Rocklin, R. E. 1977. Effect of purified alpha-fetoprotein on in vitro assays of cell-mediated immunity. *Cell. Immunol.* 30:35-42
 155. Murgita, R. A., Goidl, E. A., Kontiainen, S., Wizzell, H. 1977. Alpha-fetoprotein induces suppressor T cells in vitro. *Nature* 267:257-59
 156. Alpert, E., Dienstag, J. L., Sepersky, S., Littman, B. H., Rocklin, R. E. 1978. Immunosuppressive characteristics of human AFP: effect on tests of cell-mediated immunity and induction of human suppressor cells. *Immunol. Commun.* 7:163-85
 157. von Schoultz, B. 1974. A quantitative study of the pregnancy zone protein in the sera of pregnant and puerperal women. *Am. J. Obstet. Gynecol.* 119: 792-97
 158. Stimson, W. H. 1975. Studies on the immunosuppressive properties of a pregnancy-associated alpha-macroglobulin. *Clin. Exp. Immunol.* 25:199-206
 159. McConnell, D. J., Loeb, J. N. 1974. Kallikrein inhibitory capacity of α 2-macroglobulin subunits. *Proc. Exp. Biol. Med.* 147:891-96
 160. Tunstall, A. M., James, K. 1975. The effect of human α -macroglobulin on the restoration of humoral responsiveness in x-irradiated mice. *Clin. Exp. Immunol.* 21:173-80
 161. Cooperband, S. R., Bondevik, H., Schmid, K., Mannick, J. A. 1968. Transformation of human lymphocytes: inhibition by homologous alpha globulin. *Science* 159:1243-44
 162. Gaugas, J. M. 1974. Glycoproteins in pregnancy serum which interact with concanavalin A and may suppress lymphocyte transformation. *Transplantation* 18:538-41
 163. Bohn, H. 1971. Nachweis und charakterisierung von Schwangerschaftsproteinen in der menschlichen placenta, sowie, ihre quantitative immunologische bestimmung im serum schwangerer frauen. *Arch. Gynaekol.* 210:440-57
 164. von Schoultz, B., Stigbrand, T., Tarnvik, A. 1973. Inhibition of PHA-induced lymphocyte stimulation of the pregnancy zone protein. *FEBS Lett.* 38:23-26
 165. Kasakura, S. 1971. A factor in maternal plasma during pregnancy that suppresses the reactivity of mixed leukocyte cultures. *J. Immunol.* 107:1296-1301
 166. Stimson, W. H., Blackstock, J. C. 1976. Identification of an immunosuppressive factor in pregnancy serum. *Obstet. Gynecol.* 48:305-11
 167. Petrucco, O. M., Seamark, R. F., Holmes, K., Forbes, I. J., Sijmons, R. G. 1976. Changes in lymphocyte function during pregnancy. *Br. J. Obstet. Gynecol.* 83:245-50
 168. Breyere, E. J., Barrett, M. K. 1960. Tolerance in postpartum female mice induced by strain specific matings. *J. Natl. Cancer Inst.* 24:699-705
 169. Breyere, E. J., Barrett, M. K. 1960. Prolonged survival of skin homografts in parous female mice. *J. Natl. Cancer Inst.* 25:1405-10
 170. Kaliss, N., Dagg, N. K. 1964. Immune response engendered in mice by multiparity. *Transplantation* 2:416-25
 171. Currie, G. A. 1970. The conceptus as an allograft: immunological reactivity of the mother. *Proc. R. Soc. Med.* 63: 61-64

172. Jenkins, D. M., Hancock, K. W. 1972. Maternal unresponsiveness to paternal histocompatibility antigens in human pregnancy. *Transplantation* 13:618-19
173. Gatti, R. A., Svedmyr, E. A. J., Wigzell, H. 1974. Characterization of a serum inhibitor of MLC reactions. II. Molecular structure and dissociation of inhibition against responder and stimulator function. *Cell. Immunol.* 11:466-77
174. Gatti, R. A., Svedmyr, E. A. J., Leibold, W., Wigzell, H. 1975. Characterization of a serum inhibitor of MLC reactors. III. Specificity. *Cell. Immunol.* 15:432-51
175. Youtanankorn, V., Matangkasombut, P. 1973. Specific plasma factors blocking human maternal cell-mediated immune reaction to placental antigens. *Nature* 242:110-11
176. Pence, H., Petty, W. M., Rocklin, R. E. 1975. Suppression of maternal responsiveness to paternal antigens by maternal plasma. *J. Immunol.* 114:525-28
177. Revillard, J. P., Brochier, J., Robert, M., Bonneau, M., Traeger, J. 1976. Immunologic properties of placental eluates. *Transplant. Proc.* 8:275-79
178. Sengar, D. P. S., Opelz, G., Terrasaki, P. T. 1973. Suppression of mixed leukocyte response by plasma from hemodialysis patients. *Transplant. Proc.* 5: 641-47
179. Riggio, R. R., Saal, S. D., Stengel, K. H., Rubin, A. L. 1976. Retroplacental serum factors and passive enhancement. *Transplant. Proc.* 8:281-85
180. Riggio, R. R., Saal, S. D., Cheigh, J. S., Kim, J. J., Stubenbord, W. T., Stenzel, K. H., Rubin, A. L. 1978. Improved survival rates in presensitized recipients of kidney transplants by immunosuppression with maternal source gamma-globulin. *Lancet* 1:233-35
181. Voisin, G. A., Chaouat, G. 1974. Demonstration, nature and properties of maternal antibodies fixed on placenta and directed against paternal antigens. *J. Reprod. Fertil.* 21:89-103
182. Bonneau, M., Latour, M., Revillard, J. P., Robert, M., Traeger, J. 1973. Blocking antibodies eluted from human placenta. *Transplant. Proc.* 5:589-92
183. Souillou, J. P., Carpenter, C. B., D'A-pice, A. J. F., Strom, T. B. 1976. The role of non-classical Fc-receptor associated Ag-B antigens (Ia) in rat allograft enhancement. *J. Exp. Med.* 143:405-21
184. Chess, L., Evans, R., Humphreys, R. E., Strominger, J. L., Schlossman, S. F. 1976. Inhibition of antibody-dependent cellular cytotoxicity and Ig synthesis by antiserum prepared against a human B cell Ia-like molecule. *J. Exp. Med.* 144:113-22
185. Waldman, T. A., Broder, S. 1978. Suppressor cells in the regulation of the immune response. *Prog. Clin. Immunol.* 3:155-99
186. Cantor, H., Shen, F. W., Boyse, E. A. 1976. Separation of helper T cells from suppressor T cells expressing different Ly components. *J. Exp. Med.* 143: 1391-1403
187. Moretta, L., Ferrarini, M., Cooper, M. D. 1978. Characterization of human T cell subpopulations as defined by specific receptors for immunoglobulins. *Contemp. Top. Immunobiol.* In press
- 188a. Rocklin, R. E., Rosenthal, A. S. 1977. Evidence that human leukocyte inhibitory factor (LIF) is an esterase. *J. Immunol.* 119:249-52
- 188b. Billington, W. D. 1964. Influence of immunological dissimilarity of mother and fetus on size of placenta in mice. *Nature* 202:317-18
189. Billington, W. D. 1965. The invasiveness of transplanted mouse trophoblast and the influence of immunological factors. *J. Reprod. Fertil.* 10:343-52
190. Boshier, D. P., Moriarty, K. M. 1970. Some effects on the conceptus of prior immunological sensitization of ewes to the sire. *J. Reprod. Fertil.* 21:495-500
191. Finkel, S. I., Lilly, F. 1971. Influence of histocompatibility between mother and fetus on placental size in mice. *Nature* 234:102-3
192. Hetherington, C. M. 1973. The absence of any effect of maternal-fetal incompatibility at the H-2 and H-3 loci on pregnancy in the mouse. *J. Reprod. Fertil.* 33:135-39
193. Clarke, A. G. 1969. Factors affecting the growth of trophoblast transplanted to the testis. *J. Reprod. Fertil.* 18: 539-41
194. Koren, Z., Abrams, G., Behrman, S. G. 1968. The role of host factors in mouse trophoblastic tissue growth. *Am. J. Obstet. Gynecol.* 100:570-75
195. James, D. A. 1965. Effects of antigenic dissimilarity between mother and fetus on placental size in mice. *Nature* 205: 613-14
196. Clarke, A. G. 1971. The effects of maternal pre-immunization on pregnancy in the mouse. *J. Reprod. Fertil.* 24: 369-75
197. McLaren, A. 1975. Antigenic disparity—does it affect placental size, implantation, or population genetics? See Ref. 27, pp. 255-76

198. Beer, A. E., Billingham, R. E., Scott, J. R. 1975. Immunogenic aspects of implantation, placentation and fetoplacental growth rates. *Biol. Reprod.* 12: 176-89
199. Beer, A. E., Scott, J. R., Billingham, R. E. 1975. Histo incompatibility in maternal immunological status as determinants of fetal placental weight and litter size in rodents. *J. Exp. Med.* 142: 180-96
200. Breyere, E. J., Springer, W. W. 1969. Evidence of allograft rejection of the conceptus. *Transplant. Proc.* 1:71-75
201. Hetherington, C. M. 1971. The decidual cell reaction, placental weight, fetal weight, and placental morphology in the mouse. *J. Reprod. Fertil.* 25:417-24
202. Clarke, A. G., Hetherington, C. M. 1971. Effects of maternal preimmunization on the decidual cell reaction in mice. *Nature* 230:114-15
203. Terasaki, P. I., Mickey, M. R., Yamazaki, J. N. 1970. Maternal-fetal incompatibility. I. Incidence of HLA antibodies and possible association with congenital anomalies. *Transplantation* 9:538-43
204. Kitzmiller, J. L. 1977. Immunologic approaches to the study of preeclampsia. *Clin. Obstet. Gynecol.* 20:717-35
205. Stevenson, A. C., Davison, B. C., Say, B., Ustuoplu, S., Liya, D., Abul-Einen, M., Toppozada, H. K. 1971. Contribution of fetal/maternal incompatibility to etiology of pre-eclamptic toxemia. *Lancet* 2:1286-89
206. Scott, J. R., Beer, A. E., Stastny, P. 1976. Immunogenetic factors in pre-eclampsia and eclampsia. Erythrocyte, histocompatibility, and γ -dependent antigens. *J. Am. Med. Assoc.* 235:402-4
207. Bodmer, W. F. 1975. Discussion. See Ref. 27, p. 183.
208. Kreidl, A., Mandl, L. 1904. Über den me bergang der immonhamolysine non der frucht auf die muher. *Weinklin. Wechscht.* 17:611
209. Schinkel, P. G., Ferguson, K. A. 1953. Skin transplantation in the fetal lamb. *Aust. J. Biol. Sci.* 6:533
210. Owen, J. J. Cooper, M. D., Raft, M. C. 1974. In vitro generation of B lymphocytes in mouse fetal liver, a mammalian "bursa equivalent." *Nature* 249:361-63
211. Silverstein, A. M. 1977. Ontogeny of the immune response: A perspective. In *Development of Host Defenses*, ed. M. D. Cooper, D. Dayton, pp. 1-10. New York: Raven
212. Zinkernagel, R. M., Callahan, G. N. Klein, J., Deunert, G. 1978. Cell surface changes in allogantigen activated T lymphocytes. *Nature* 271:251-53
213. Osoba, D. 1965. Immune reactivity in mice thymectomized soon after birth: normal responses after pregnancy. *Science* 147:298-99
214. Cooper, M. D., Dayton, D. H. eds. 1977. *Development of Host Defenses*. New York: Raven
215. Jones, W. R. 1976. Fetal and neonatal immunology. See Ref. 2, pp. 127-67
216. Linscott, W. D. 1970. Effect of cell surface antigen density on immunological enhancement. *Nature* 228:824-27
217. Bansal, S. C. 1974. Cytotoxicity and blocking factors in neoplasia. In *Progress in Immunology II*, ed. L. Brent, J. Holborow, pp. 394-97, Vol. 3. Amsterdam: North-Holland
218. Tung, K. 1974. Immune complexes in the renal glomerulus during normal pregnancy. *J. Immunol.* 112:186-200

GROWTH CONSIDERATIONS IN THE RADIATION THERAPY OF CHILDREN WITH CANCER

♦7331

*Philip S. Littman, M.D. and Giulio J. D'Angio, M.D.*¹

Department of Radiation Therapy, University of Pennsylvania
School of Medicine, Philadelphia, PA 19104

INTRODUCTION

Radiation therapy (RT) plays a major role in the treatment of children with cancer. As in adults, RT is used to cure and to palliate, but modern pediatric cancer care also commonly entails the use of combined modality therapy. Local control of the primary tumor and its regional extensions is achieved by surgery, often assisted by postoperative RT, while remote microscopic malignant cell deposits are destroyed by anticancer chemotherapeutic agents or by RT. Coordinated triple modality care has proven very successful, and has widened its scope to include some conditions and stages of disease previously considered unsuitable for irradiation. For example, cranial or craniospinal axis RT is now commonly added to chemotherapy regimens for the management of children with leukemia to eradicate occult foci of meningeal disease. In the realm of "solid tumors," less radical surgical procedures are being used than heretofore in some instances; RT and chemotherapy assist in achieving local control. A gratifying rise in survival rates has resulted from such coordinated programs. More than 50% of children with acute leukemia can now be expected to survive more than five years (1), and the two-year survival of Wilms' tumor patients is now better than 80% (2).

Many children with these hitherto highly lethal conditions can be expected to survive for long periods, and are therefore more likely to develop late complications (3). There is another important factor: RT effects are

¹Supported in part by grant CA 11796 from the National Institutes of Health, and by the Fondation Dressler, Geneva, Switzerland.

most pronounced in rapidly dividing cells such as those found in malignant tumors. However, the tissues and organs in childhood are also undergoing rapid growth and development, and thus they, too, are more subject to damage than their counterparts in adults. Finally, some of the anticancer drugs used in pediatrics are radiation enhancers, and augment RT effects not only in malignant cells but in normal tissues as well (4, 5). For all these reasons, it is important to understand the late adversities of successful cancer treatment so that these complications can be prevented whenever possible, or minimized if they cannot be avoided (Figure 1).

Modern radiation therapy is given with high energy or supervoltage machines, by which is meant those delivering radiation with energies greater than one million electron volts. The radiations may be produced by radioactive sources such as cobalt-60, or by electrical generators (most commonly, linear accelerators). Prior to 1960, radiation therapy was delivered with so-called orthovoltage equipment. Energies were in the 200–300-

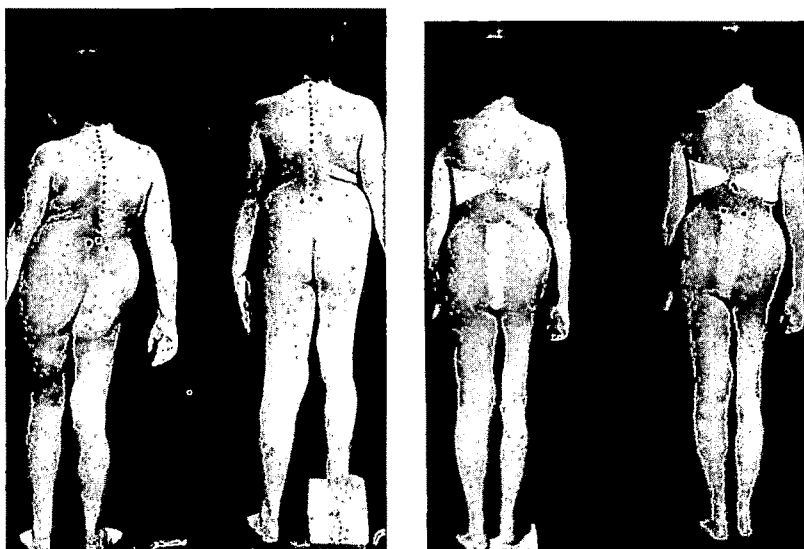


Figure 1 Left: A dose of under 1000 rads was delivered to the right knee of this girl when she was four years of age because of a soft tissue tumor near the knee. The leg is short and the soft tissues underdeveloped seven years later. Note the lift required under the foot to stabilize the pelvis.

Right: The distal femoral and proximal tibial epiphyses were arrested on the untreated side. The legs were more nearly equal in length at the time this photograph was taken, and became almost identical when full growth was completed. (From Pearson, D., D'Angio, G. J. 1975. Radiation therapy. In *Cancer in Children*, ed. H. J. G. Bloom, J. Lemerle, M. K. Neidhardt, P. A. Voûte, p. 33. Berlin: Springer)

kilovolt range. The disadvantage of the older machines was the decreased depth of penetration, which caused more pronounced reactions in superficial soft tissues than are caused by higher energy units. Another important consideration, especially in pediatrics, is the fact that kilovolt x rays are absorbed more readily in bone than are those of megavoltage energy. Thus, for a similar dose delivered to a deep-seated tumor, less damage is incurred in surrounding bone and soft tissues when megavoltage techniques are used. Much of the experience recorded concerning disruption of normal growth is based on results obtained in the orthovoltage era. The observations nonetheless are an important source of clinical radiobiologic information pertaining to pediatric radiation therapy.

Both clinical observations and animal experiments indicate that the effects of radiation therapy on growth depend on the anatomical area treated, the dose, and the age of the subject. The most obvious changes are those found in the bony skeleton, so the discussion that follows is largely concerned with bone growth.

RT effects can be either direct through interaction of the ionizing beam and the affected cell, or indirect through changes in adjoining structures or through disruption of normal endocrine gland function.

DIRECT EFFECTS

Animal Experiments

One of the first experiments to demonstrate that ionizing radiation damaged dividing cells was performed by Perthes in 1903 (6). He showed that irradiation of the wings of a day-old chick resulted in shortened wings. It has subsequently been demonstrated by numerous investigators that growing bone is very sensitive to ionizing radiations. Microscopic changes in the epiphyseal cartilage can be seen in young rabbits after single exposures of 200 rad, and the threshold dose for measurable inhibition of bone growth is in the 400–500-rad range (7). Gall and co-workers (8) found that the degree of growth retardation was proportional to the dose delivered to the epiphyseal plates of rat long bones. Engle (9) showed that growth of long bones in rabbits was inversely related to the age of the animals at the time of irradiation. Engle (10) and Arkin et al (11) demonstrated that asymmetrical deformities could be produced by irradiating one half of the vertebral bodies of animals, which then developed scolioses.

These experiments were refined in a series of studies by Rubin et al (12), who showed differential effects when the several segments of young animal long bones were irradiated. They used strontium-90, a radioactive nuclide taken up in the epiphyseal plate where cartilage and osteoid are calcified. The resultant uniform irradiation led to symmetrical stunting of the extrem-

ities. External beam irradiation of the metaphysis resulted in failure of tubulation and consequent bowing of the extremity. Diaphyseal RT disrupted normal periosteal bone formation so that the bones became narrowed in their midportion.

These and other laboratory experiments in growing animals (13) substantiate that a variety of different types of deformity can be produced depending on the bone and the portion of bone treated, and that the effects are directly related to dose and inversely to age.

Clinical Observations

Bone and soft tissue deformity in children secondary to radiation therapy has been recorded in many reports that document changes in both tubular and in flat bones.

Neuhauser et al (14) in a classic study first showed the relationship between dose and age at the time of treatment. They reported their results in 34 children who had had radiation therapy for abdominal tumors, and in whom the portal had included the lumbar vertebrae (Figure 2). Doses of 1000 rad or less at any age produced little if any evidence of growth disturbance. The most pronounced changes were seen with doses in excess of 2000 rads in children under two years of age. Neuhauser and his col-

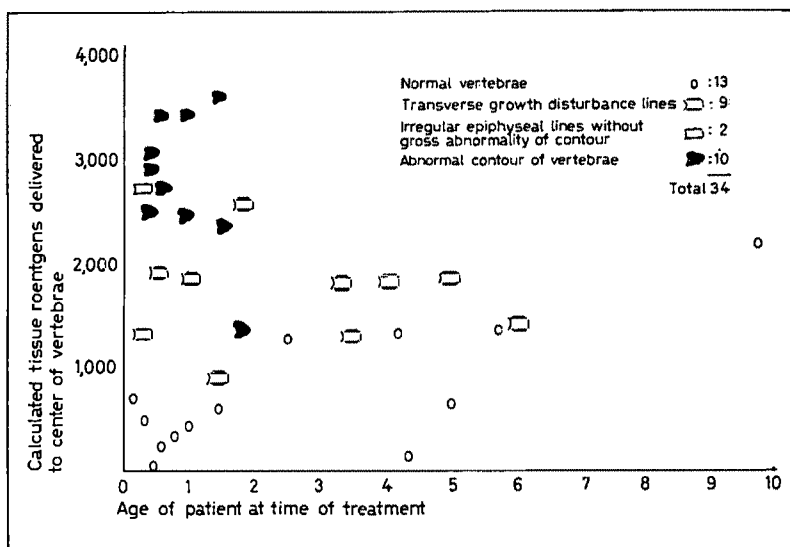


Figure 2 Postradiation vertebral body roentgenographic appearances according to dose and age as described by Neuhauser et al (14).

leagues understood that other factors might be responsible for shortening of the vertebral body height and for the scolioses that are sometimes seen after treatment of children with cancer. These factors include the presence of associated soft tissue anomalies such as in patients with neurofibromatosis or in those who have posterior mediastinal neuroblastomas that affect the developing rib cage.

These observations, and the fact that the disruption of normal bone growth can also become pronounced at the time of the adolescent growth spurt have been documented by many other clinical investigators. That there are two periods during childhood that are most sensitive to bone growth disruption was nicely shown by Probert and co-workers (15, 16). They measured the sitting height of 22 patients who received axial irradiation for Hodgkin's disease or medulloblastoma, and who were given doses of 3500–4000 rads. All the patients had roentgenographic changes in the vertebral bodies, and 16 of the 22 children showed retardation of spinal growth. The sitting height of 10 was shorter by more than two standard deviations from the normal mean for the age, and major growth changes were more likely to occur in those under six years of age or in those experiencing the adolescent growth spurt. The changes appeared to represent a temporary cessation of growth, with a resumption of the normal rate thereafter and a resultant net loss that did not worsen with time. This same sequence has been noted both in experimental systems and in other clinical observations. The nontubular bones elsewhere share in the expression of post radiation bone growth disruption. Inclusion of the iliac apophysis in the field used for post operative RT of the flank for Wilms' tumor, for example, leads to underdevelopment of the hemipelvis on the treated side (12, 17, 18). Undergrowth of the bones of the face also follows therapeutic doses of radiation, and in the case of the mandible or the maxilla, eruption of the teeth is either delayed or truncated with consequent dental problems (19).

The most pronounced effects are seen after treatment of the growth centers at the knee (Figure 1). Metaphyseal sclerosis and fraying and epiphyseal widening supervene; the changes superficially resemble those found in rickets (20).

INDIRECT EFFECTS

Local

Reduction in muscle mass may follow bone shortening occasioned by RT (13). Conversely, indirect effects in bone can be produced by RT-associated fibrosis and reduction of muscle length, which can lead to secondary bending of bone through a "bow string" mechanism (13).

Abscopal

Normal growth is mediated by proper hormonal balances, and these can be disrupted by radiation of the relevant endocrine sources. It has become apparent in the last decade or so that the hypothalamic-pituitary axis and the thyroid gland, structures hitherto thought to be relatively radioresistant, actually can be affected adversely by relatively modest doses of RT given to children. Pituitary dwarfism has been reported after high radiation doses were given to the pituitary region; for example, a 12-year-old girl received 9300 rads to the base of the brain for nasopharyngeal carcinoma and developed panhypopituitarism (21). Similar observations have been recorded by several other authors (22–28). It is now known, however, that modest doses of RT also can result in depression of normal growth hormone levels. This has been documented, for example, in children with acute leukemia who receive cranial irradiation for suppression of meningeal leukemia. Shalet and his colleagues (27) found an impaired growth hormone response to hypoglycemia and arginine stimulation in four of 15 patients followed 2½ to 5 years after treatment. Eight had received 2500 rads in 10 fractions, and seven were given 2400 rads in 20 treatment days. All those with demonstrable growth hormone deficiency had received the more intensive course, and none of those receiving the more protracted regimen had abnormal levels. Thus, the changes appear to be related to radiation therapy, and can be minimized if not avoided entirely by suitable modifications of the time-dose-fractionation schedules adopted for RT. That the growth disturbance is secondary to disruption of growth hormone is established by the increase in growth velocity apparent after treatment with growth hormone.

RT to the hypothalamic-pituitary region may cause other abnormalities. Perry-Keene et al (26) reported evidence of abnormalities of both the TSH and ACTH levels, as did Samaan and colleagues (24). Thus, the thyroid is indirectly subject to suppression, and normal thyroid function in children is essential for normal growth and development. RT to the neck may result more directly in decreased thyroxin production. This is noted most often in patients receiving neck irradiation for Hodgkin's disease, when a preceding lymphangiogram appears to make such thyroid suppression much more likely (28). It is thought that the small amounts of iodine released from the lymphangiographic contrast material in some way enhances the suppressive effects of RT. It is important to recognize this potential complication, and to institute replacement therapy promptly.

RT to the pelvis can result in the suppression of ovarian function after relatively modest doses (29). This effect is most pronounced in girls at or near the menarche. Suppression of ovarian function is associated with a

delay in the closure of the epiphyses; therefore, following RT castration these girls continue to grow after the usual age when growth stops. Replacement therapy ordinarily is given these individuals, however, so that there is little likelihood that abnormal growth patterns will be seen in many girls. The secretory cells of the testis are relatively radioresistant, and hypogonadism after RT of the scrotal contents is rare. Spermatogonia, on the other hand, are very radiosensitive, so that oligospermia or aspermia follows relatively modest radiation doses to the adolescent or postadolescent testis (30). Such doses may be received concomitant to pelvic radiation for Hodgkin's disease, for example.

OTHER ORGANS AND STRUCTURES

Virtually every organ or structure in a child can be severely disrupted in its normal growth patterns after therapeutic doses of RT. Multiple examples could be given, such as underdevelopment of the aorta and its major branches (31, 32) or agenesis of the female breast (19) (Figure 3). The liver (25) and the kidney (34, 35) are particularly susceptible when they are stimulated to hypertrophy after unilateral nephrectomy or partial hepatec-



Figure 3 Irradiation of the left upper chest and shoulder girdle for an intrathoracic neuroblastoma was given when this girl was three months of age. The dose was approximately 2000 rads. Underdevelopment of the bony and soft tissue structures resulted, including agenesis of the left breast.

tomy (33). The brain is likely to be especially sensitive before the fifth year, during which time the central nervous system structures are still developing (36, 37).

COMBINED CHEMO- AND RADIATION THERAPY

Several of the anticancer chemotherapeutic agents used in pediatric oncology are radiation enhancers. These include actinomycin-D (AMD), adriamycin, and high doses of methotrexate and of cyclophosphamide (5, 38). Animal experiments demonstrate bone growth disruption that is more pronounced when RT is combined with AMD in animal systems. Murine studies in our laboratories are relevant. The rear legs, including the knee, of 4-weeks-old mice were given single doses of 1000 or more rad with and without 0.18 mg/kg body weight of actinomycin D intraperitoneally. The growth of the legs of the irradiated animals, five per group, was compared with that of untreated controls and of those given AMD alone. Growth inhibition was expressed as a percentage of growth attained by untreated controls nine weeks after treatment. Mice given AMD alone grew normally. Those receiving RT at 1000, 1250, and 1500 rad showed 5, 25, and 50% suppression of expected growth, respectively; while those given RT + AMD showed 50% of their expected growth at all three doses. Higher RT doses produced even more pronounced deformity; only 20% of normal growth followed 2000 rad. The result was not changed when AMD was added. These data suggest that AMD effects are most evident at relatively low RT doses; in this case, 0.18 mg/kg body weight of AMD added to 1000 rad produced an effect equivalent to that after 1500 rad alone.

Clinically, combinations of multiple radiation-enhancing drugs have been found to produce severe changes within bone and soft tissues, severely disrupting the integrity of the part receiving combined therapy (38). The drugs have an additional effect; namely, they reactivate latent radiation responses. Skin reactions, for example, can be made to "flare" with each cycle of chemotherapy given after the completion of RT. Extreme caution, therefore, is necessary whenever combined programs using multiple agent chemotherapy and RT are instituted, especially when known radiation enhancers are part of the program.

Because the anticancer agents are effective in suppressing microscopic deposits of disease, it has been one of the theses of pediatric oncology recently that chemotherapy contributes to local control as well as to the ablation of disease at more remote sites. This hypothesis is being tested by some of the national studies dealing with pediatric tumors, such as the National Wilms' Tumor Study and the Intergroup Rhabdomyosarcoma Study. Initial results appear to confirm those expectations (2, 39). Studies are now planned in the case of Wilms' tumor to determine whether RT

doses can be reduced or omitted entirely with safety in more advanced stages of the disease and thus make the complications of combined therapy less severe.

A VIEW TO THE FUTURE

Radiation therapy is an effective modality in the management of children with cancer (40). The associated damage, however, has long caused concern among pediatric oncologists and radiation therapists. Many investigations have been undertaken in order to define more precisely the indications for RT in the management of children with a variety of neoplastic processes, rather than using it routinely as had been the practice in several of the entities under discussion. For instance, the first National Wilms' Tumor Study sought to determine whether the prevailing practice of delivering routine postoperative radiation therapy was needed in all patients with nephroblastoma, or whether it could be safely omitted in children with tumors confined to the kidney and totally removed surgically. It was found that the survival was equally good in children under two years of age with tumors of that type who did and did not receive postoperative radiation therapy (2). A similar question is being asked in the Intergroup Rhabdomyosarcoma Study, in which patients are randomized between regimens that do and do not include postoperative RT when excision of the primary tumor appears complete (39).

Continuing studies are underway for these and other cancers of childhood. Precise definition of the RT doses to be given in order to secure a beneficial effect are needed, and what those doses should be when combined with radiation-enhancing chemotherapeutic agents.

The interplay of drugs and radiation therapy has other implications as well. Ionizing radiation is an oncogenic agent, and both benign and malignant tumors are known to occur in irradiated areas (41, 42). Investigations in progress seek to determine whether combined therapy leads to a change in the frequency of radiation-associated tumors. There are at least two clinical observations that bear on this point. AMD was found to reduce the risk of radiation-associated tumors in long-term survivors treated for childhood cancer (43). On the other hand, recent evidence suggests that intensive chemotherapeutic regimens combined with RT increases the likelihood of acute myelogenous leukemia developing in patients with Hodgkin's disease (44, 45).

There is, therefore, every reason to explore fully the role of RT in modern pediatric oncology to refine the indications for its use. This very useful but potentially damaging modality should be used only in those instances where radiation therapy has been demonstrated to be of distinct value in improving survival rates.

Literature Cited

1. Pinkel, D. 1976. Treatment of acute leukemia. *Pediatr. Clin. North Am.* 23:117-30
2. D'Angio, G. J., Evans, A. E., Breslow, N., Beckwith, B., Bishop, H. et al. 1976. The treatment of Wilms' tumor: results of the National Wilms' Tumor Study. *Cancer* 38:633-46
3. D'Angio, G. J. 1977. Late adversities of treatment in long-term survivors of childhood cancer. *Proc. Am. Cancer Soc. 2nd Natl. Conf. Human Values Cancer*, pp. 59-63, Chicago, Ill.
4. D'Angio, G. J. 1962. Clinical and biologic studies of actinomycin-D and roentgen irradiation. *Am. J. Roentgenol.* 87:106-9
5. Phillips, T. L., Fu, K. K. 1976. Quantification of combined radiation therapy and chemotherapy effects on critical normal tissue. *Cancer* 37:1186-1200
6. Perthes, G. 1903. Über den Einfluss der Roentgenstrahlen auf epitheliale gewebe, insbesondere auf das carcinom. *Arch. Klin. Chir.* 79:955-1000
7. Bisgard, J. D., Hunt, H. B. 1936. Influence of roentgen rays and radium on epiphyseal growth of long bones. *Radiology* 26:56-68
8. Gall, E. A., Lingley, J. R., Hilcken, J. A. 1940. Comparative experimental studies of 200 and 1,000 kV roentgen rays. I. The biological effect on the epiphysis of the albino rat. *Am. J. Pathol.* 16:605-18
9. Engel, D. 1938. An experimental study of the action of radium on developing bones. *Br. J. Radiol.* 11:779-803
10. Engel, D. 1939. Experiments on the production of spinal deformities by radium. Part I. *Am. J. Roentgenol.* 42:217-34
11. Arkin, A., Simon, N., Siffert, R. 1948. Asymmetrical suppression of vertebral epiphyseal growth with ionizing radiation. *Proc. Soc. Exp. Biol. Med.* 69:171-73
12. Rubin, P., Duthie, R. B., Young, L. W. 1962. The significance of scoliosis in post-irradiated Wilms' tumor and neuroblastoma. *Radiology* 79:539-59
13. D'Angio, G. J., Jung, J., Wright, K. A., Cohen, J. 1964. Disturbance in bone growth produced by 3.5 mev electron irradiation of young rabbit limbs. *Am. J. Roentgenol. Radium Ther.* 91:1132-37
14. Neuhauser, E. B. D., Wittenborg, M. H., Berman, L. Z., Cohen, J. 1952. Irradiation effects of roentgen therapy on the growing spine. *Radiology* 59:637-50
15. Probert, J. C., Parker, B. R. 1975. The effects of radiation therapy on bone growth. *Radiology* 114:155-62
16. Probert, J. C., Parker, B. R., Kaplan, H. S. 1973. Growth retardation in children after megavoltage irradiation of the spine. *Cancer* 32:634-39
17. Riseborough, E. J., Grabias, S. L., Burton, R. I., Jaffe, N. 1976. Skeletal alterations following irradiation for Wilms' tumor. *J. Bone Joint Surg.* 58-A:526-36
18. Katzman, H., Waugh, T., and Berdon, W. 1969. Skeletal changes following irradiation of childhood tumors. *J. Bone Joint Surg.* 51-A:825-42
19. Dawson, W. B. 1968. Growth impairment following radiotherapy in childhood. *Clin. Radiol.* 19:241-56
20. Rutherford, H., Dodd, G. 1974. Complications of radiation therapy: growing bone. *Semin. Roentgenol.* 9:15-27
21. Tan, B. C., Kumaratnum, N. 1966. Hypopituitary dwarfism following radiotherapy for nasopharyngeal carcinoma. *Clin. Radiol.* 17:302-4
22. Harrop, J. S., Davies, T. J., Capra, L. G., Marks, V. 1976. Hypothalamic-pituitary function following successful treatment of intracranial tumors. *Clin. Endocrinol.* 5:313-21
23. Richards, G. E., Wara, W. M., Grombush, M. M., Kaplan, S. L., Sheline, G. E., Conte, F. A. 1976. Delayed onset of hypopituitarism: Sequelae of therapeutic irradiation of central nervous system, eye, and middle ear tumors. *J. Pediatr.* 89:553-59
24. Samaan, N. A., Bakdash, M. M., Cade-rao, J. B., Cangir, A., Jesse, R. H., Ballantyne, A. J. 1975. Hypopituitarism after external irradiation. *Ann. Intern. Med.* 83:771-77
25. Shalet, S. M., Beardwell, C. G., Morris-Jones, P. H., Pearson, D. 1975. Pituitary function after treatment of intracranial tumours in children. *Lancet* 2:104-7
26. Perry-Keene, D. A., Connelly, J. F., Young, R. A., Wettenthal, H. N. B., Martin, F. I. R. 1976. Hypothalamic hypopituitarism following external radiotherapy for tumors distant from the adenohypophysis. *Clin. Endocrinol.* 5:373-80
27. Shalet, S. M., Beardwell, C. G., Morris-Jones, P. H. 1976. The effect of varying doses of cerebral irradiation on growth hormone production in childhood. *Clin. Endocrinol.* 5:287-90
28. Fuks, Z., Glatstein, E., Marsa, G. W., Bagshaw, M. A., Kaplan, H. S. 1976.

- Long-term effects of external radiation on the pituitary and thyroid glands. *Cancer* 37:1152-61
29. Shalet, S. M., Beardwell, C. G., Morris-Jones, P. H., Pearson, D., Orrell, D. H. 1976. Ovarian failure following abdominal irradiation in childhood. *Br. J. Cancer* 33:655-58
 30. Lushbaugh, C. C., Ricks, R. C. 1972. Some cytokinetic and histopathologic considerations of irradiated male and female gonadal tissue. *Front. Radiat. Ther. Oncol.* 6:228-48
 31. Rotman, M., Seidenberg, B., Rubin, I. 1969. Aortic arch syndrome secondary to radiation in childhood. *Arch. Intern. Med.* 124:87-96
 32. Lee, D. H., Sapire, D., Markowitz, R., Gruskin, A. 1976. Radiation injury to abdominal aorta and iliac artery sustained in infancy. *S. Afr. Med. J.* 50:658-60
 33. Tefft, M., Mitus, A., Das, L., Vawter, G. F., Filler, R. M. 1970. Irradiation of the liver in children: review of experience in the acute and chronic phases and in the intact normal and partially resected. *Am. J. Roentgenol.* 108: 368-85
 34. Levitt, W. M., Oram, S. 1956. Irradiation-induced malignant hypertension: cured by nephrectomy. *Br. Med. J.* 2:910-12
 35. Staab, G., Tegtmeier, C., Constable, W. 1976. Radiation-induced renovascular hypertension. *Am. J. Roentgenol.* 126:634-37
 36. Zeman, W. 1968. The effects of atomic radiations. In *Pathology of the Nervous System*, ed. J. Minckler, Ch. 65, Vol. 1, pp. 864-939. New York: McGraw-Hill
 37. Yakovlev P. I. 1967. The myelogenetic cycles of regional maturation of the brain. In *Regional Development of the Brain in Early Life*, ed. A. Minkowski, pp. 3-71. Oxford: Blackwell
 38. Tefft, J., Lattin, P. B., Jereb, B., Cham, W., Ghavimi, F. et al. 1976. Acute and late effects of normal tissues following combined chemo- and radiotherapy for childhood rhabdomyosarcoma and Ewing's sarcoma. *Cancer* 37:1201-13
 39. Maurer, H. M., Moon, T., Donaldson, M. H., Fernandez, C., Gehan, E. A., Hammond, D., Hays, D. M., Lawrence, W. Jr., Newton, W., Ragab, A., Raney, R. B., Soule, E. H., Sutow, W. W., Tefft, M. 1977. The Intergroup Rhabdomyosarcoma Study: A preliminary report. *Cancer* 40:2015-26
 40. D'Angio, G. J. 1978. Radiation therapy for solid tumors. In *Pediatric Surgery*, ed. M. R. Ravitch, K. Welch et al. Chicago: Year Book Medical. 3rd ed. In press
 41. Fry, R. J. M. 1977. Radiation carcinogenesis. *Int. J. Radiat. Oncol. Biol. Phys.* 3:219-26
 42. Meadows, A. T., D'Angio, G. J., Evans, A. E., Harris, C. C., Miller, R. W., Miké, V. 1975. Oncogenesis and other late effects of cancer treatment in children. Report of a single hospital study. *Radiology* 114:175-80
 43. D'Angio, G. J., Meadows, A. T., Miké, V., Harris, C., Evans, A. E. et al. 1976. Decreased risk of radiation-associated second malignant neoplasms in actinomycin-D treated patients. *Cancer* 37:1177-85
 44. Canellos, G., Arseneau, J. C., DeVita, V. T., Whang-Peng, J., Johnson, R. 1975. Second malignancies complicating Hodgkin's disease in remission. *Lancet* 1:947-49
 45. Coleman, C. N., Williams, C. J., Flint, A., Glatstein, E. J., Rosenberg, S. A., Kaplan, H. S. 1977. Hematologic neoplasia in patients treated for Hodgkin's disease. *N. Engl. J. Med.* 297:1249-52

PRENATAL SEX HORMONES ◆7332

AND THE DEVELOPING BRAIN:

Effects on Psychosexual Differentiation and Cognitive Function

*Anke A. Ehrhardt, Ph.D. and Heino F. L.
Meyer-Bahlburg, Dr. rer. nat.*

New York State Psychiatric Institute and Department of Psychiatry, College of
Physicians and Surgeons of Columbia University, New York, New York 10032

INTRODUCTION

Animal experimental research has demonstrated crucial effects of sex hormones on the developing central nervous system (CNS). Deprivation of testosterone in the male animal during his species-specific critical time of brain differentiation will result in a female pattern of sex-dimorphic behavior. In contrast, a female rat given testosterone shortly after birth will, as an adult, perform more like a male in the tests for male and female sexual behavior. This principle holds true even though mammalian behavior is not completely sex dimorphic but is a matter of degree and relative frequency of behavior components that are more typical for one sex or the other (see 1).

Considering the evidence from animal studies, it is a reasonable assumption that sex hormones may also have an influence on human brain differentiation and thus, play a role in psychosexual differentiation. Any such effects would be of immediate clinical relevance for the management of patients with a problem of intersexuality and for the evaluation of pregnancy drugs. Over the last decade or so, a number of studies have focused on prenatal sex hormones and behavior. However, as becomes evident in this review, our current knowledge is still fragmentary.

Four behavior areas have been studied:

1. Gender identity: the primary identification with one sex or the other.
2. Gender role or sex-dimorphic behavior: all those aspects of behavior in which normal girls and boys differ from each other in our culture and at this time in history.
3. Sexual orientation: erotic attraction to one sex or the other, i.e. homo-, hetero-, and bisexuality.
4. Intelligence and cognitive sex differences: strengths and weaknesses of mental abilities as related to sex in normal subjects.

One common experimental approach is the manipulation of prenatal or neonatal hormones by castration of the male animal or by exogenous administration of androgens to the female. Because of obvious ethical limitations, researchers of human behavior are typically unable to design ideal experimental studies, but rather depend on spontaneous endocrine abnormalities or unusual prenatal hormone histories due to drug treatment of the pregnant mother. Available studies can be divided by type of hormone into those that have attempted to assess prenatal androgen effects in females and males and those that have looked at the possible implications of abnormal levels of progestogens and estrogens for postnatal behavior. We review the evidence from studies on patients with endocrine syndromes first.

ENDOCRINE SYNDROMES

The study of patients with documented endocrine conditions has been particularly useful in clarifying the role of prenatal androgen in psychosexual differentiation. Therefore, we restrict our review to endocrine syndromes including an excess or lack of androgen during fetal development.

Fetal Androgenization in Genetic Females and Males

In this context, patients with congenital adrenal hyperplasia (CAH) are of particular interest. The primary defect of CAH is a genetic one, transmitted as an autosomal recessive, which prevents the adrenal cortices from synthesizing cortisol. Instead, they release an excess of adrenal androgens from fetal life onward. As a result, the external genitalia are masculinized if the fetus is a genetic female, while genetic males are born with normal male external genitalia. Postnatally, the condition can be corrected by replacement therapy with corticosteroids that suppress the adrenal androgens to normal levels if medical management is optimal. In girls, external genitalia can be surgically feminized in the first weeks of life. If properly regulated on cortisone or prednisone, pubertal development occurs spontaneously at the normal age for both females and males. In adulthood, sexual functioning and fertility does not appear to be impaired (2).

For the study of the effects of prenatal androgens, those children who are treated with cortisone beginning in infancy are the most suitable subjects. If treatment is instituted later in life, androgenization continues with resulting masculinization of the body. In those cases, it is difficult to separate prenatal from postnatal hormonal effects. Over the last decade, several studies have been reported on the behavior of children with the early-treated syndrome of CAH.

PSYCHOSEXUAL DIFFERENTIATION Gender identity typically agrees with the sex of rearing of the child, i.e. females with CAH firmly identify as girls and women in spite of prenatal androgenization and masculinization of the external genitalia, provided sex assignment is clearly female from early childhood (2, 3). Genetic males with CAH are always raised as boys and identify as such (4, 5). In the past, genetic females with a severe degree of genital virilization were sometimes by mistake raised as boys. In such cases, gender identity typically is male (6). However, if confusion about a child's sex of rearing persists beyond early childhood, in the mind of parents and others, there is a risk of ambiguity in the child's gender identity formation, which sometimes results in sex reassignment later in life (7).

While prenatal androgenization does not appear to affect gender identity formation directly, it seems to influence gender role behavior. Studies with three different designs have been reported: (a) comparison of patients with matched normal controls (8), (b) comparison of patients with unaffected siblings (4), and (c) comparison of patients with a clinical contrast group (9). The two studies using a control group or a sibling comparison design included genetic females who were early treated and early surgically corrected. In both samples, the prenatally androgenized females differed significantly from the controls in several behavior aspects: They typically demonstrated a cluster of high energy expenditure (intense active outdoor play, more male friends, and long-term identification as a tomboy by self and others) and a cluster of decreased parenting behavior (low interest in doll play and baby care, less interest in role rehearsal of wife and mother versus having a career). The characteristic behavior was not episodic, but rather a long-term pattern of childhood. As such, it was not considered abnormal or pathological but rather fell within one area of the spectrum for acceptable female behavior in our culture. Most patients of both studies were well controlled by corticosteroids. Therefore, the behavior pattern seemed to be mostly attributable to prenatal effects of excess adrenal androgens, in spite of the fact that it is sometimes difficult to institute optimal medical regulation.

In one other study, similar behavior aspects were assessed in girls with CAH (10). The authors were unable to confirm earlier findings. However, differences in medical management of the sample employed and psychologic

assessment procedures may account for the different results. For instance, the authors included patients with significantly reduced height age. This may have been a result of later onset of treatment or a too high dosage of corticosteroids. In contrast, the patients in the other two studies fell within the normal range of body height. Therefore, medical management may have been quite dissimilar for patients of the McGuire et al (10) study and may in turn have been an intervening variable as far as behavior development is concerned.

The evidence so far suggests that excessive prenatal androgens play a role, albeit a subtle rather than a powerful one, in psychosexual differentiation of females with CAH.

We also reported findings on the behavior of boys with CAH (11). They differed from a male sibling control group only in that they exhibited higher levels of energy expenditure in play and sports. This observation strengthens the assumption that prenatal androgens may be related to the temperamental trait of intense active outdoor play.

Sexual orientation has been assessed in a total of 56 genetic female patients with CAH (4, 12–14), three of whom were raised as males. The majority of females were described as heterosexual, some were bisexual, none exclusively homosexual; this was true even for the patients who did not have adequate medical treatment before late adolescence and adulthood. Whether the incidence of bisexuality in the female patients was any higher than in endocrinologically normal women is questionable, because control groups were not employed. The three patients raised as males were heterosexual with respect to their sex of rearing. Two different studies (4, 5) reported sexual orientation in genetic male patients with CAH as heterosexual in all cases.

The evidence so far supports the theory that the decisive variable for sexual orientation of patients exposed to excessive prenatal androgenization is their sex of rearing rather than their fetal history. Even merely a facilitory neuroendocrine predisposition by high levels of prenatal androgens for a homosexual orientation in females has not yet been established, although the current evidence does not rule it out either (see 15).

INTELLIGENCE General level of mental functioning has been reported as elevated in a high percentage of female and male patients with CAH (16–19). However, more recently it has been found that siblings and parents of CAH patients also tend to be above normal in full IQ (16, 17). The elevation of intelligence in total families with one or more CAH children may be due to a socioeconomic bias of patients who come to specialty clinics, although socioeconomic level of CAH families has been reported as not different from the general population or other clinic patient populations (16, 20). Alterna-

tively the recessive genetic trait may be somehow linked to another trait favoring postnatal intellectual development.

Apart from general intelligence level, no particular pattern of cognitive strengths or weaknesses has emerged, except possibly for relatively inferior number abilities compared to overall level of mental functioning (16, 19, 21).

Fetal Nonandrogenization in Genetic Males

The human clinical syndrome that most closely approximates the antithesis of fetal androgenization of the genetic female is the syndrome of androgen insensitivity (testicular feminization) of the genetic male. The tissues of the body are insensitive to androgen, so that the differentiation of the external genitalia is female, except for a blind vagina which often requires surgical lengthening in or after middle teenage to permit satisfactory sexual intercourse. Gonadectomy is typically performed in early childhood or adolescence. Breast development is normal in response to the estrogen produced by the gonads or to estrogen replacement therapy after gonadectomy. Menstruation and fertility are missing. Sex of rearing is female.

PSYCHOSEXUAL DIFFERENTIATION Gender identity is typically female in these patients (22). Gender role behavior has been systematically assessed in adult patients by their current adult behavior and in retrospect by their childhood behavior (23). In the majority of the cases, their pattern of sex-dimorphic behavior was stereotypically feminine with a preference for being a wife with no outside job, for enjoying homecraft (including having dreams and fantasies of raising their family), for having played primarily with dolls and other girls' toys in childhood, and for having a genuine interest in infant care. Sexual orientation is usually heterosexual in congruence with their sex of rearing as female (22, 24).

In these cases, the presumptive prenatal hormone influence is nonmasculinizing, which is in agreement with the postnatal behavior pattern of femininity. Since the postnatal rearing history is also typically that of a female, one cannot separate the causative role of endocrine and social factors.

Fetal nonandrogenization can also occur in a less complete form with partial masculinization at birth. Dependent on the degree of male development of the genitalia, sex of rearing is male or female. If surgical correction to achieve an erectile penis in adulthood is doubtful, patients have been found to do better in the female role (2). In either case, psychosexual differentiation typically follows the sex of rearing. These cases demonstrate that the sex chromosomes, gonads, and sex hormones by themselves do not dictate postnatal gender identification. The evidence for the powerful effects

of consistent early social-environmental conditioning on gender identity formation is even stronger when one considers that children with both ambiguity of their sex organs at birth and subsequent exposure to confusion by pubertal hormones and secondary sex characteristics discordant with their sex of rearing still identify firmly with their assigned gender (25).

A syndrome of decreased dihydrotestosterone due to 5 α -reductase deficiency was described in a group of patients in the Dominican Republic (26). Of particular interest were eight genetic males who were assigned to the female sex at birth, and grew up with small phalli, bifid scrotums, and urogenital sinuses. The testes were in the urogenital canals or labial-scrotal folds. At puberty, under the influence of testosterone a muscular male habitus developed with some growth of the phallus and scrotum, voice change, no gynecomastia, and decreased facial and body hair development (27). All but one of the eight affected individuals with this clinical history assumed the male role in spite of their rearing as females (28). While this report seems to contradict the theory that gender identity follows the sex of rearing, one has to be cautious in interpreting these observations, since crucial information on parent-child relationship, the specific cultural interpretation of masculine and feminine roles, and the patients' own emotions, attitudes, and body image is lacking. Considering the fact that they were known as "guevedoces" (penis at 12) to their community, it seems likely that these individuals were not unambiguously raised female, but rather had a special status as hermaphrodites with the knowledge of an impending sex change at puberty.

INTELLIGENCE In patients with the syndrome of androgen insensitivity, the level of intelligence does not differ from the norm, and the cognitive pattern of strengths and weaknesses is unremarkable except for a significant and consistent, but modest, tendency toward verbal rather than space-form abilities (29).

HORMONE ADMINISTRATION DURING PREGNANCY

The effectiveness of progestogens and estrogens prescribed for problem pregnancies is usually assessed in terms of rates of spontaneous abortion, premature births, infant mortality, and morbidity; and the evaluation of side effects focuses on the well-being of the mother and on congenital malformations of the newborn. Since it is well documented that problem pregnancies put the surviving offspring at risk for neurological and intellectual deficits, the inclusion of such variables in treatment-outcome research seems very important. Observations in the late 1950s of increased rates of

pseudohermaphroditism in offspring from hormone-treated pregnancies, together with the animal data of the effects of such hormones on brain and behavior, also call for an evaluation of behavioral sequelae. However, pertinent investigations are rare and often wanting in methodology because the necessary long-term follow-up studies pose tremendous logistical problems. Thus, the information available is very fragmentary, especially in view of the great number of pregnancy hormones and their combinations in past and current use.

Progestogens

PSYCHOSEXUAL DIFFERENTIATION Extensive animal research has demonstrated both androgenic and antiandrogenic effects of various progestogens on genital differentiation, gonadotropin regulation, and sex-dimorphic behavior. Dorfman & Kincl (30) were the first to suggest that progesterone may protect the developing hypothalamic-pituitary-gonadal system of the fetus against possibly damaging effects of increases in androgen and/or estrogen concentrations. Later findings of sex differences in fetal and/or neonatal progesterone levels—with the females having the higher levels—in the rhesus monkey and the rat led to general models of an interaction of progesterone with other sex steroids in both male and female fetuses (31, 32). In these theories, progesterone is assumed to protect the female brain from the masculinizing effects of androgens and estrogens.

In humans, genital masculinization of a significant minority of female newborns has been found after administration of certain C-19-derived progestogens such as 17 α -ethinyltestosterone or 17 α -ethinyl-19-nortestosterone (33). Individual cases of masculinization after prenatal exposure to exogenous progesterone (e.g. 33) or to derivatives such as medroxyprogesterone acetate (e.g. 34) have also been described, but a causal relationship has not been established because of the rarity of such patients. Antiandrogenic effects of exogenous progestogens during prenatal development are suggested by the observation in a few human newborn males that hypospadias is associated with various progestogens administered during pregnancy (e.g. 35). It is obvious, however, that the overwhelming majority of male and female offspring from progestogen-treated pregnancies show normal external genitalia. For instance, a major prospective study, the Collaborative Perinatal Project, revealed only 13 children with disorders of the genitourinary system, including hypospadias, among 1144 mother-child pairs who were exposed to progestogens (as progestational agents or oral contraceptives) during the first four months of pregnancy. This was practically identical with the expected rate of 11:1 for offspring not exposed to drugs (36). Although genital malformations indicate significant abnormali-

ties in hormone availability to or utilization by peripheral target tissues, genital findings cannot directly be generalized to the CNS systems that regulate hormone-dependent sex-dimorphic behaviors, since the biochemical mechanisms involved in the sexual differentiation of the brain are known to differ from those in the periphery.

The first preliminary report on the effects of exogenous progestogen treatment during pregnancy on psychosexual development was published by Ehrhardt & Money (37). Comparisons with pair-matched controls involving detailed blind rating procedures were reported subsequently (2, 38). The index subjects were ten girls, aged 3–14 years, who had been prenatally exposed to C-19–derived synthetic progestins; nine were born with somewhat masculinized external genitalia and some of these were surgically feminized. On the basis of detailed behavioral interview data, nine index subjects were rated to be tomboys. The progestin-treated girls rated significantly higher than their controls in athletic interests and skills, in preference for male versus female playmates, for utilitarian and functional rather than traditionally feminine clothing, for toy cars, guns, etc versus dolls, and in priority of career over marriage. Although these girls exhibited long-term tomboyism, their gender identity was feminine. There is also no evidence that such females, after puberty, develop a homosexual orientation (39).

Effects of exogenous progesterone were investigated in a blind interview study by Zussman et al (40, 41) in 12 females and 18 males, aged 16–19 years, compared to 29 matched controls. Progesterone had been administered during pregnancy for the relief of precursor symptoms of preeclamptic toxemia. Progesterone exposure in boys correlated negatively with physical activity level in childhood as well as with heterosexual activity level in adolescence. Progesterone exposure in girls was negatively correlated with reports of tomboyism and positively with traditionally feminine activities in childhood. The preliminary reports do not give sufficient information on the subjects' background variables, so that it is difficult to decide whether and to what extent the behavioral differences described are due to differences in pregnancy conditions or socioeconomic level rather than progesterone treatment. Effects on gender identity or sexual orientation were not reported. Medroxyprogesterone acetate (MPA) treatment was the focus of a double-blind interview study by Meyer-Bahlburg and Ehrhardt on 13 boys (42) and 15 girls (43), aged 8–14 years, with closely pair-matched controls. Whereas the MPA-treated boys did not differ significantly from the controls with respect to a variety of sex-dimorphic behaviors, the hormone-exposed girls showed a lower incidence than controls of being labeled a tomboy during childhood and a more frequent preference for feminine clothing styles. Gender identity seemed not to be affected.

Combining the available data on the behavioral effects of progestogen treatment, the results reported seem compatible with the theory of hormone-influenced differentiation of sex-dimorphic behavior: C-19 progestogens seem to have shifted the behavior of girls to the masculine side while progesterone and MPA showed opposite effects. However, the effects noted in the progesterone and MPA studies are weak, even with relatively high dosage as in the sample described by Zussman et al (41), and it seems possible that they were determined by factors other than hormones, for instance by the pregnancy condition. It is noteworthy in this connection that recent studies could not establish any sex difference in progesterone levels of umbilical-cord blood (44, 45). So the validity of the theory that progesterone has an androgen-protective effect on the developing brain has not yet been established for humans.

INTELLIGENCE Effects of exogenous prenatal progestogens on subsequent cognitive functioning have been studied by several teams. In their above-mentioned report on ten girls with prenatal progestogen treatment, Ehrhardt & Money (37) also presented data on intelligence. As a group, those girls had unusually high IQs but it is not clear to what extent this was a reflection of the elevated socioeconomic status of their parents since no control group was used for the IQ data.

A series of progesterone studies was reported by Dalton (46) in England. She had available children whose mothers were treated with high dosages of progesterone for preeclamptic toxemia of pregnancy, a condition that may lead to intellectual impairment. A pilot study of 32 children, aged 6–13 years, showed that the offspring of these mothers received intelligence ratings of "above average" by their head teachers slightly more frequently than the children of a matched normal control group. Dalton then investigated the offspring of mothers with toxemic pregnancies who had been randomly assigned to the progesterone treatment or to a nonprogesterone control condition. At age one year, the 29 progesterone-exposed children appeared somewhat more advanced in standing and possibly in walking, but not with regard to other milestones. Dalton also compared a sample of 29 progesterone-exposed children, aged 9–10 years, to normal controls and to untreated toxemic controls; teacher ratings suggested elevated school performance significantly more frequently for the hormone-exposed subjects than for the controls. In a very recent follow-up study of 34 progesterone-exposed children from toxemic pregnancies, Dalton (47) noted that a higher proportion of exposed children than normal or untreated toxemic controls left school with advanced grades and entered a university.

It appears that the four studies by Dalton are not independent but involve largely overlapping samples of progesterone-exposed children and controls.



Unfortunately, the reports lack detail in data presentation, statistical analysis, and background information so that it is difficult to judge to what degree possible confounding variables were controlled. Dalton's samples were re-studied with psychological tests by Zussman et al (40, 41). These authors noted elevations in subtests of the Differential Aptitude Test—statistically significant only for numerical ability—although their preliminary report does not yet allow conclusive interpretations. As in the earlier studies by Dalton, the major problem is the control-group selection.

Meyer-Bahlburg & Ehrhardt (48) studied the effects of prenatal MPA treatment on 28 female and male subjects in comparison to 28 closely pair-matched controls. There were no differences in verbal, performance, or full IQs (WISC-R), Cohen's verbal, number, and perceptual factors, or individual subtests. Although this study appears to have applied considerably stricter criteria for control-group matching, the results do not necessarily refute Dalton's data, since she used another hormone in considerably higher dosage on a differently defined sample.

Recent endocrine reports, however, raise further doubts as to the possible effectiveness of exogenous progesterone in preventing CNS damage resulting from high-risk pregnancies. Progesterone has been found to be within normal limits or slightly elevated in such high-risk pregnancy conditions as preeclampsia or diabetes (see references in 49). Also, Escarcena et al (50) demonstrated that, at term, about 1% of the maternally circulating progesterone crosses the placenta and thus constitutes 10% or less of the hormone in fetal circulation. If these figures apply to earlier stages of pregnancy as well, even the highest treatment dosage reported (300 mg daily i.m.; 46), which is roughly equivalent to the maternal production rate of progesterone, would not contribute more than about 10% of the fetal circulation; and less extreme treatment regimens would remain far below this figure. The same argument applies to the presumptive positive effects of progesterone on brain development in normal pregnancy. A recent attempt (51) to replicate such progesterone effects in normal pregnancy in rats was unsuccessful.

Estrogens and Progestogens

PSYCHOSEXUAL DIFFERENTIATION Exogenous estrogens, both steroidal and nonsteroidal, have been shown to exert paradoxical effects on genital morphology and behavior in subhuman mammals, i.e. to masculinize females and to demasculinize males. Analogous paradoxical modifications of the genitalia have been described in a few scattered cases of human infants after exposure to diethylstilbestrol (DES; summarized in 52) but,

again, the rarity of such cases does not permit stringent causal conclusions. (The more recently described cancerigenic and other effects of prenatal DES on the genitourinary tract in male and female offspring is not considered here, because they do not fit the changes typical of pseudohermaphroditism.) So far, no investigation has been reported on behavior effects of prenatal estrogen treatment alone in man. While our incomplete knowledge of the behavior effects of the various progestational and estrogenic agents individually makes a prediction of their interactive effects prohibitive, data are available from several studies that focused on estrogen-progestogen combinations in human subjects.

Yalom et al (53) studied a sample of 20 males, 16–17 years old, who had been exposed in utero to DES plus progesterone during medication for diabetic pregnancy. In comparison to untreated controls, the subjects were found to have somewhat decreased aggression/assertion and decreased athletic coordination (by observational ratings) with tendencies to decreased overall masculine interests and decreased heterosexual experience. Increased gender identity problems or homosexuality were not noted. The same authors studied a second group of 20 boys at age six years, who showed decreased assertion and decreased athletic ability (by teacher ratings) after they had been exposed in utero to estradiol valerate plus hydroxyprogesterone acetate, again during treatment for diabetic pregnancy; two experimental subjects had hypospadias. Unfortunately, neither study employed an adequate pair-matched control group of children with the same intrauterine history but without hormone treatment; thus, we do not know if the observed effects resulted from the hormone treatment or the medical condition of the mother.

Reinisch & Karow (54, 55) studied by questionnaire 71 children (26 boys, 45 girls), aged 5–17 years, who had been exposed prenatally to various combinations of synthetic progestogens and estrogens, with untreated siblings as controls. The data on the Cattell Personality Questionnaires showed significant treatment effects: Subjects who were mainly exposed to progestogens were characterized as more independent, sensitive, self-assured, individualistic, and self-sufficient, while subjects who were predominantly exposed to estrogens were more group oriented and group dependent. Unfortunately, this study does not permit any conclusions regarding sex-dimorphic behavior, since the latter was not assessed by the questionnaires used and the study subjects were not analyzed by sex.

INTELLIGENCE Reinisch & Karow (55) also provided data on cognitive functioning. They divided their sample of 71 hormone-exposed subjects into those who were exposed predominantly to estrogens, to progestogens, or to

an intermediate combination of both. Neither the subgroups nor the total sample showed any significant differences in Wechsler IQs from untreated sibling controls.

CONCLUSIONS

The empirical evidence strongly suggests effects of prenatal hormones, albeit quite limited, on sex-dimorphic behaviors such as physical energy expenditure, childhood rehearsal of parenting, and related behaviors. Best established is the role of prenatal androgens, whereas the evidence for an antiandrogenic action of prenatal progesterone and related compounds seems insufficient as yet. A role of prenatal hormones for sexual orientation has not been demonstrated. The development of gender identity depends on the sex of rearing rather than on chromosomes, gonads, and prenatal hormones. Gender identity may become ambiguous if there are inconsistencies in prenatal sex typing. Concerning the effects of prenatal sex hormones on general intelligence or on sex differences in mental abilities, the available data are inconclusive.

ACKNOWLEDGMENT

The work by the authors was supported in part by a grant from the Spencer Foundation.

Literature Cited

1. McEwen, B. S. 1976. Interactions between hormones and nerve tissues. *Sci. Am.* 235:48-58
2. Money, J., Ehrhardt, A. A. 1972. *Man and Woman, Boy and Girl: The Differentiation and Dimorphism of Gender Identity from Conception to Maturity*. Baltimore, Md: Johns Hopkins Univ. Press. 311 pp.
3. Money, J., Hampson, J. G., Hampson, J. L. 1955. An examination of some basic sexual concepts: The evidence of human hermaphroditism. *Bull. Johns Hopkins Hosp.* 97:301-19
4. Ehrhardt, A. A., Baker, S. W. 1974. Fetal androgen, human CNS differentiation and behavior sex differences. In *Sex Differences in Behavior*, ed. R. C. Friedman, R. M. Richart, R. L. Vande Wiele, pp. 53-76. New York: Wiley. 495 pp.
5. Money, J., Alexander, D. 1969. Psychosexual development and absence of homosexuality in males with precocious puberty: Review of 18 cases. *J. Nerv. Ment. Dis.* 148:111-23
6. Money, J., Ehrhardt, A. A. 1968. Prenatal hormonal exposure: possible effects on behavior in man. In *Endocrinology and Human Behaviour*, ed. R. P. Michael, pp. 32-48. London: Oxford Univ. Press. 349 pp.
7. Money, J. 1968. Psychologic approach to psychosexual misidentity with elective mutism: Sex reassignment in two cases of hyperadrenocortical hermaphroditism. *Clin. Pediatr.* 7:331-39
8. Ehrhardt, A. A., Epstein, R., Money, J. 1968. Fetal androgens and female gender identity in the early-treated adrenogenital syndrome. *Johns Hopkins Med. J.* 122:160-67
9. Ehrhardt, A. A., Evers, K., Money, J. 1968. Influence of androgen and some aspects of sexually dimorphic behavior in women with the late-treated adrenogenital syndrome. *Johns Hopkins Med. J.* 123:115-22
10. McGuire, L. S., Ryan, K. O., Omenn, G. S. 1975. Congenital adrenal hyperplasia. II. Cognitive and behavioral studies. *Behav. Genet.* 5:175-88

11. Ehrhardt, A. A. 1975. Prenatal hormone exposure and psychosexual differentiation. In *Topics in Psychoendocrinology*, ed. E. J. Sachar, pp. 67-82. New York: Grune & Stratton. 182 pp.
12. Ehrhardt, A. A. 1973. Maternalism in fetal hormonal and related syndromes. In *Contemporary Sexual Behavior. Critical Issues in the 1970's*, ed. J. Zubin, J. Money, pp. 99-115. Baltimore: Johns Hopkins Univ. Press. 468 pp.
13. Money, J., Daléry, J. 1976. Iatrogenic homosexuality: gender identity in seven 46,XX chromosomal females with hyperadrenocortical hermaphroditism born with a penis, three reared as boys, four reared as girls. *J. Homosex.* 1:357-71
14. Money, J., Schwartz, M. 1977. Dating, romantic and non-romantic friendships, and sexuality in 17 early-treated adrenogenital females, aged 16-25. In *Congenital Adrenal Hyperplasia*, ed. P. A. Lee, L. P. Plotnick, A. A. Kowarski, C. J. Migeon, pp. 419-31. Baltimore: Univ. Park Press. 532 pp.
15. Meyer-Bahlburg, H. F. L. 1979. Sex hormones and female homosexuality: a documentation. *Arch. Sex. Behav.* 8: In press
16. Baker, S. W., Ehrhardt, A. A. 1974. Prenatal androgens, intelligence and cognitive sex differences. See Ref. 4, pp. 33-51
17. McGuire, L. S., Omenn, G. S. 1975. Congenital adrenal hyperplasia. I. Family studies of IQ. *Behav. Genet.* 5: 165-73
18. Money, J., Lewis, V. 1966. IQ, genetics and accelerated growth: Adrenogenital syndrome. *Bull. Johns Hopkins Hosp.* 118:365-73
19. Perlman, S. M. 1973. Cognitive abilities of children with hormone abnormalities: screening by psychoeducational tests. *J. Learn. Disabil.* 6:26-34
20. Money, J., Lewis, V. G., Ehrhardt, A. A., Drash, P. W. 1967. IQ impairment and elevation in endocrine and related cytogenetic disorders. In *Psychopathology of Mental Development*, ed. J. Zubin, pp. 22-27. New York: Grune & Stratton. 658 pp.
21. Lewis, V. G., Money, J., Epstein, R. 1968. Concordance of verbal and non-verbal ability in the adrenogenital syndrome. *Johns Hopkins Med. J.* 122:192-95
22. Masica, D. N., Money, J., Ehrhardt, A. A. 1971. Fetal feminization and female gender identity in the testicular feminizing syndrome of androgen insensitivity. *Arch. Sex. Behav.* 1:131-42
23. Money, J., Ehrhardt, A. A., Masica, D. N. 1968. Fetal feminization induced by androgen insensitivity in the testicular feminizing syndrome: Effect on marriage and maternalism. *Johns Hopkins Med. J.* 123:105-14
24. Lewis, V. G. 1977. Androgen insensitivity syndrome: Erotic component of gender identity in nine women. In *Progress in Sexology*, ed. R. Gemme, C. C. Wheeler, pp. 61-62. New York: Plenum. 615 pp.
25. Ehrhardt, A. A. 1978. Psychosexual adjustment in adolescence in patients with congenital abnormalities of their sex organs. In *Proceedings of the Birth Defects Institute Symposium VII on Genetic Mechanisms of Sexual Development*. New York: Academic. In press
26. Imperato-McGinley, J., Guerrero, L., Gautier, T., Peterson, R. E. 1974. Steroid 5 α -reductase deficiency in man: an inherited form of male pseudohermaphroditism. *Science* 186:1213-15
27. Peterson, R. E., Imperato-McGinley, J., Gautier, T., Sturla, E. 1977. Male pseudohermaphroditism due to steroid 5 α -reductase deficiency. *Am. J. Med.* 62:170-91
28. Imperato-McGinley, J., Peterson, R. E. 1976. Male pseudohermaphroditism: the complexities of male phenotypic development. *Am. J. Med.* 61:251-72
29. Masica, D. N., Money, J., Ehrhardt, A. A., Lewis, V. G. 1969. IQ, fetal sex hormones, and cognitive patterns: Studies in the testicular feminizing syndrome of androgen insensitivity. *Johns Hopkins Med. J.* 124:34-43
30. Dorfman, R. I., Kincl, F. A. 1963. A new function of progesterone. IV. Reun. Anu., Soc. Mex. Nutr. Endocrinol., Dec., p. 171. (Cited in: Dorfman, R. I. 1967. The antiestrogenic and antiandrogenic activities of progesterone in the defense of a normal fetus. *Anat. Rec.* 157:547-58)
31. Resko, J. A. 1974. The relationship between fetal hormones and the differentiation of the central nervous system in primates. In *Reproductive Behavior*, ed. W. Montagna, W. A. Sadler, pp. 211-22. New York, London: Plenum. 376 pp.
32. Shapiro, B. H., Goldman, A. S., Bongiovanni, A. M., Marino, J. M. 1976. Neonatal progesterone and feminine sexual development. *Nature* 264: 795-96

33. Wilkins, L. 1960. Masculinization of female fetus due to use of orally given progestins. *J. Am. Med. Assoc.* 172: 1028-32
34. Burstein, R., Wasserman, H. C. 1964. The effect of provera on the fetus. *Obstet. Gynecol.* 23:931-34
35. Aarskog, D. 1971. Intersex conditions masquerading as simple hypospadias. *Birth Defects: Original Article Series*, Vol. VII, No. 6, pp. 122-30
36. Heinonen, O., Slone, D., Shapiro, S. 1977. *Birth Defects and Drugs in Pregnancy*. Littleton, Mass: Publ. Sci. Group. 516 pp.
37. Ehrhardt, A. A., Money, J. 1967. Progesterin-induced hermaphroditism: IQ and psychosexual identity in a study of ten girls. *J. Sex Res.* 3:83-100
38. Ehrhardt, A. A. 1969. *Zur Wirkung fötaler Hormone auf Intelligenz und geschlechtsspezifisches Verhalten*. PhD thesis. Univ. Düsseldorf, 151 pp. (In German)
39. Mathews, D., Money, J. 1978. Progesterin-induced hermaphroditism: a follow-up report. In *East. Conf. Reprod. Behav., Madison, Wisc., May 31-June 3, 1978. Program Abstr.*, p. 62 (Abstr.)
40. Zussman, J. U., Zussman, P. P., Dalton, K. 1975. *Post-pubertal effects of prenatal administration of progesterone*. Presented at Meet. Soc. Res. Child Dev., Denver
41. Zussman, J. U., Zussman, P. P., Dalton, K. 1977. *Effects of prenatal progesterone on adolescent cognitive and social development*. Presented at 3rd Ann. Meet. Int. Acad. Sex Res., Bloomington, Ind.
42. Meyer-Bahlburg, H. F. L., Grisanti, G. C., Ehrhardt, A. A. 1977. Prenatal effects of sex hormones on human male behavior: medroxyprogesterone acetate (MPA). *Psychoneuroendocrinology* 2: 383-90
43. Ehrhardt, A. A., Grisanti, G. C., Meyer-Bahlburg, H. F. L. 1977. Prenatal exposure to medroxyprogesterone acetate (MPA) in girls. *Psychoneuroendocrinology* 2:391-98.
44. Tulchinsky, D., Okada, D. M. 1975. Hormones in human pregnancy. IV. Plasma progesterone. *Am. J. Obstet. Gynecol.* 121:293-99
45. Dawood, M. Y., Helmkamp, F. 1977. Human umbilical arterial and venous progesterone concentrations. *Obstet. Gynecol.* 50:450-53
46. Dalton, K. 1968. Ante-natal progesterone and intelligence. *Br. J. Psychiatry* 114:1377-82
47. Dalton, K. 1976. Prenatal progesterone and educational attainments. *Br. J. Psychiatry* 129:438-42
48. Meyer-Bahlburg, H. F. L., Ehrhardt, A. A. 1977. Effects of prenatal hormone treatment on mental abilities. See Ref. 24, pp. 85-92
49. Vande Wiele, R. L., Husami, N., Dyrenfurth, I. 1976. Progestogens in pregnancy. In *Diabetes and other Endocrine Disorders during Pregnancy and in the Newborn*, ed. M. I. New, R. H. Fifer, Jr., pp. 179-203. New York: Liss. 262 pp.
50. Escarcena, L., Clark, H., Gurpide, E. 1978. Contribution of maternal circulation to blood-borne progesterone in the fetus. I. Studies on human subjects. *Am. J. Obstet. Gynecol.* 130:462-65
51. Coyle, I. R., Anker, R., Cragg, B. 1976. Behavioral, biochemical and histological effects of prenatal administration of progesterone in the rat. *Pharmacol. Biochem. Behav.* 5:587-90
52. Bongiovanni, A. M. 1976. Androgens, estrogens, and the fetus. See Ref. 49, pp. 173-78
53. Yalom, I. D., Green, R., Fisk, N. 1973. Prenatal exposure to female hormones. *Arch. Gen. Psychiatry* 28:554-61
54. Reinisch, J. M. 1977. Prenatal exposure of human fetuses to synthetic progestin and estrogen affects personality. *Nature* 266:561-62
55. Reinisch, J. M., Karow, W. G. 1977. Prenatal exposure to synthetic progestins and estrogens: effects on human development. *Arch. Sex. Behav.* 6:257-88

TREATMENT OF PLASMA CELL MYELOMA

♦7333

*Daniel E. Bergsagel, M.D., D. Phil.*¹

Department of Medicine, University of Toronto, and Ontario Cancer Institute,
Toronto, Canada M4X 1K9

INTRODUCTION

The investigation of plasma cell neoplasms during the past two decades has revealed much about the biology of the tumors, the structure and functions of the immunoglobulins (M-proteins) they produce, the prognostic factors, the clinical staging, and the response to treatment. Attempts to improve the effectiveness of treatment have been disappointing.

In this review I consider the prognostic factors and assess the effectiveness of various drug combinations in the treatment of myeloma patients. The failure of intensive chemotherapy to increase the duration of remissions or survival, and the absence of a correlation between cell-kill and survival suggests that the beneficial effects of treatment may not be directed against the neoplastic plasma cells. Evidence presented in the final section suggests that a growth control mechanism may be operative in some patients with plasma cell neoplasms.

PROGNOSTIC FACTORS

Many of the clinical trials designed to evaluate the effectiveness of treatment have now been analyzed to identify the important clinical features that influence prognosis (1-4). General features of the disease, the immunoglobulin class of M-protein, clinical manifestations related to the total body myeloma cell number, and renal function have all been found to influence prognosis.

¹Supported by Grant 143, Ontario Cancer Treatment and Research Foundation.

General Features

The performance status of a patient summarizes the effects of pain, anorexia, weight loss, anemia, renal failure, and other manifestations of the disease. It is not surprising to find that myeloma patients who present with a good performance status have the best survival prognosis (1).

A history of weight loss has not been examined critically as a prognostic factor in myeloma, though weight loss is frequently reported by myeloma patients. It is probably just as important a prognostic factor in myeloma as it is in the malignant lymphomas.

An unexplained fever frequently develops during the acute terminal phase of myeloma (5). Survival following the appearance of this fever is short, for all of the 17 patients included in this report died within 1-9 months (median 3 months) after the fever was recognized.

Immunoglobulin Class of M-Protein

Neoplasms arising from cells that produce different immunoglobulins might be expected to express different biologic properties, such as preferences for the growth of neoplastic cells in certain tissues, and differences in growth rate. Indeed, plasma cell tumors producing different classes of immunoglobulins do have different clinical manifestations. IgM-producing plasma cell neoplasms frequently cause lymphadenopathy and splenomegaly, but lytic bone lesions are uncommon. In contrast, myelomas producing other types of immunoglobulins usually cause lytic bone lesions, while splenomegaly and lymphadenopathy are unusual.

The type of M-protein produced by the plasma cell neoplasm also influences survival (6). Those producing IgM are reported (7) to have the best survival (median 50 months). The median survival of IgG myeloma patients is about 29 months (8, 9), for IgA about 21 months (8, 9), and 30 months for those producing only kappa light chains (9, 10). The shortest median survivals reported are 9 months for IgD myeloma (11), and 10 months for those producing only lambda light chains (9, 10). A retrospective analysis of the clinical manifestations and response to therapy of 52 myeloma patients producing kappa, and 45 producing lambda, light chains failed to identify a reason for the shorter survival of patients with lambda light-chain disease (10). No differences in the degree of anemia, uremia, hypercalcemia, hypoalbuminemia, Bence Jones proteinuria, the severity of lytic skeletal disease, the frequency of amyloidosis, or the therapeutic response to alkylating agents were found in patients with kappa and lambda light-chain disease. Similarly, it was not possible to explain the shorter survival of patients with lambda light-chain disease on the basis of the stage of disease when treatment was started. Thus, the explanation for the shorter survival of patients with lambda light-chain disease remains to be found.

A review of the factors affecting the survival of patients treated by the Cancer and Acute Leukemia Group B revealed that patients who excrete light chains in the urine have a worse prognosis than those who do not. The survival prognosis for those excreting type kappa was significantly better than for those excreting type lambda, if the blood urea nitrogen (BUN) was less than 30 mg/dl. The type of light chain excreted did not affect the prognosis of patients with a BUN greater than 30 mg/dl (8). The poor prognosis for IgD myeloma patients is probably related to the amount and type of light chain excreted, for 90% of these patients excrete large amounts of lambda light chains, and 67% present with a BUN of greater than 30 mg/dl (11).

Clinical Manifestations Related to Total Myeloma Cell Number

The total myeloma cell number can be estimated by dividing the total body M-protein synthesis rate by the synthesis rate of single myeloma cells in vitro (12). Durie & Salmon (3) estimated the myeloma cell mass of 71 myeloma patients and made a search for laboratory and clinical parameters that would predict the total myeloma cell number. They found that the myeloma cell mass could be accurately predicted from: (a) extent of bone lesions, (b) hemoglobin level, (c) serum calcium level, and (d) the M-protein levels in serum and urine. A clinical staging system (3) that uses these four parameters divides patients into those with low (Stage I), intermediate (Stage II), and high (Stage III) myeloma cell numbers. This staging system is useful for stratifying patients for randomization in clinical trials.

Blood Urea Nitrogen

Many investigators have recognized that an elevated BUN is an adverse prognostic factor in myeloma patients (1-4, 8). In the first Medical Research Council (MRC) myelomatosis trial, the initial blood urea concentration proved such a strong determinant of prognosis that its influence had to be taken into account in assessing whether other prognostic factors acted independently (2). The median survival of 37 months for patients with a BUN less than 20 mg/dl, decreased to 20 months for those with a BUN of 20-39 mg/dl and to 2 months if the BUN was greater than 40 mg/dl (2). The adverse effect of light-chain proteinuria is closely linked with the occurrence of uremia; the latter occurs more frequently in patients who excrete large amounts of light-chain protein. This association is in keeping with the view that light-chain proteinuria is harmful to renal function. The MRC group (2) found that the correlation of proteinuria with prognosis disappeared after adjustment for blood urea concentration. Others found that for patients with an initial BUN of less than 30 mg/dl, the excretion of lambda light chains shortened the survival of myeloma patients produc-

ing an IgG M-protein (8), or light chains only (10), whereas the excretion of kappa light chains did not adversely affect survival.

Renal function is the major factor affecting prognosis that appears to act independently of the total body myeloma cell number (3). Thus, the prognosis for a patient who presents with a BUN of greater than 40 mg/dl is equally bad for patients with Stages I, II, and III disease.

THE EFFECTIVENESS OF TREATMENT

Good objective responses to systemic chemotherapy were rarely observed in myeloma patients prior to the use of alkylating agents. The demonstration that melphalan relieved bone pain, caused weight gain, decreased the serum M-protein (which occasionally disappeared from the serum electrophoresis pattern), decreased proteinuria to normal, shrank plasmacytomas, and occasionally healed lytic skeletal lesions was greeted with great enthusiasm (13–15). The fact that a few patients appeared to achieve complete remissions stirred the hope that it might be possible to eliminate all neoplastic plasma cells by chemotherapy. A more careful evaluation of the cell-kill achieved by treatment with an alkylating agent and prednisone revealed that good remissions were achieved with a reduction in the myeloma cell mass of one half to one order of magnitude and that a reduction by more than two orders of magnitude was unusual (16, 17).

Many chemotherapeutic agents have been tested against both murine (18) and human plasma cell tumors. Most of the alkylating agents tested appear to be about as active as melphalan. Unfortunately, few nonalkylating agents are active against plasma cell neoplasms. A few, such as procarbazine (19, 20), vincristine (21), adriamycin (22, 23), and *cis*-platinum (24), reportedly show some activity against mouse or human myeloma.

All attempts to improve the response and survival of myeloma patients by using combinations of chemotherapeutic agents have yielded disappointing results. The comparison of the effectiveness of drug combinations in Table 1 has been restricted to an analysis of the results of cooperative groups that use the Southwest Oncology Group (SWOG) criteria for the evaluation of response (25).

The addition of prednisone (P) to intermittent courses of melphalan (M) doubled the response rate but had little influence on survival. The improved response rate could result from effects of P that are not directed at myeloma cells. For example, a P-stimulated increase in the catabolic rate of the M-protein (29), or a decrease in bone resorption as a result of P blocking the effects of osteoclast-activating factor (30), could increase the response rate without changing the myeloma cell mass. Even though P may not have a strong antineoplastic action against myeloma cells, its nonspecific effects

Table 1 Effectiveness of combination chemotherapy for plasma cell myeloma

Drugs ^a	Group	No. responsive ^b		Response rate (%)	Survival ^c median (mo.)	Reference
		No. evaluable				
M	SWOG	13/54		24	18	25
MP	SWOG	67/139		48	21	25
	NCI-C	40/100		40	28	26 ^e
MP + Proc.	SWOG	120/205		59	23	25
MAP	SWOG	31/67		46	21-26 ^d	27
CAP	SWOG	23/59		39	21-26 ^d	27
MCP	SWOG	35/74		47	21-26 ^d	27
MCBP	SWOG	34/70		49	~34	27
	NCI-C					
	alternating	28/91		31	27	26 ^e
	concurrent	39/83		47	30	26 ^e
VMP + Proc.	SWOG	64/127		50	34	27, 28
VMCP	SWOG	48/77		62	~34	27
VCAP	SWOG	42/74		57	~34	27

^a Abbreviations used are: M = melphalan, P = prednisone, Proc. = procarbazine, A = adriamycin, C = cyclophosphamide, B = carmustine, V = vincristine.

^b Responsive as defined by SWOG (25), i.e. $\geq 75\%$ regression.

^c Survival from the onset of treatment.

^d The survival of these groups of patients is reported to be within this range.

^e Updated.

are so useful that it is now included in most therapeutic drug trials for plasma cell myeloma.

Combining procarbazine with melphalan and prednisone (MP) appeared to increase the response rate but, again, had little effect on survival. Further experience suggests that procarbazine contributes little in the management of myeloma, and it has been dropped from subsequent SWOG myeloma trials.

The addition of adriamycin (A) to MP or cyclophosphamide (C) and P, did not improve the response rate or survival.

The discovery that mouse (31) and human (32) plasma cell tumors, which were resistant to one alkylating agent, still responded to another drug of this class, indicated that different alkylating agents are not necessarily cross resistant. Furthermore, combinations of alkylating agents were shown to be synergistic in the treatment of two murine tumors (33, 34). These observations stimulated clinical trials of combinations of alkylating agents in the treatment of plasma cell myeloma. The SWOG found that combinations containing MC, or MC and carmustine (B), did not significantly change the response rate or survival. Similarly, the National Cancer Institute of Canada (NCI-C) Clinical Trials group found no advantage for the MCBP

combination, administered either alternately or concurrently, over MP alone. In contrast, the Cancer and Acute Leukemia Group B (CALGB) found that MCBP significantly improved the response rate in several parameters over that obtained with MP alone, but did not improve the survival of the patients (35). The CALGB study of MCBP differs from those of the SWOG and NCI-C in that all of the alkylating agents in the combination were administered intravenously; the results were compared with the oral administration of M to the MP group. Since the absorption of M may be erratic (36, 37), the improved response rate could be the result of the intravenous administration of M to the MCBP group.

Studies of the labeling index of myeloma marrow plasma cells before and after treatment demonstrated a marked rise in the index after treatment with alkylating agents (21, 38). This observation led to the suggestion that an increase in the growth fraction after treatment maintained the myeloma cell mass constant after an initial reduction of one to two orders of magnitude. Thus, a trial of cycle-specific agents in these patients was initiated. In a preliminary trial, vincristine (V) caused a modest further decrease in myeloma cell number (21), and was selected for further study in drug combinations. The SWOG was encouraged by the modest improvement in response rate and survival of patients treated with combinations containing V (27, 29). However, as noted in Table 1, combinations that do not contain V reportedly produce similar high response rates (e.g. MP + Procarbazine) and survival (e.g. MCBP). The M-2 protocol for myeloma patients, employed at Memorial Hospital in New York, combines VMCBP; this combination reportedly produces a higher response rate and longer survival than MP (39). However, the trial of the M-2 protocol was not a controlled study; patients treated with VMCBP were compared with a historical control group treated with MP. The fact that only one of 46 patients died during the first 10 months on VMCBP suggests that poor-risk patients may not have been entered in this trial, for most myeloma survival curves show a loss of 20–25% within 10 months (26, 27).

CELL-KILL ACHIEVED AND REMISSION DURATION AND SURVIVAL

If the improved survival of myeloma patients who respond to treatment results from the effect of that treatment on myeloma cells, one would expect to find a correlation between the estimated cell-kill and survival. The estimated myeloma cell-kill following 12 courses of treatment varied from $10^{-0.3}$ to $10^{-6.4}$, with survivals ranging from three months to more than 10 years: No correlation was demonstrated between cell-kill and survival (17).

The results of treating a patient with total body irradiation are also consistent with this view (17). The radiation sensitivity of myeloma cells (40) suggested that a dose of 225 rads would reduce the myeloma cell number by a factor of ten and result in a remission of about three years. The administration of this dose of irradiation to one patient caused the serum M-protein to disappear, and lytic skeletal lesions to heal; this excellent remission has now persisted for more than eight years. It is not possible to explain the duration of this remission on the basis of the myeloma cell-kill expected from the dose of irradiation that was used.

The failure of maintenance therapy to prolong the duration of remissions or survival (28, 41) provides additional evidence that these myeloma-response parameters are not determined by the tumor cell-kill achieved by treatment.

INCREASED GROWTH RATE OF MYELOMA CELLS

The myeloma growth rate, as reflected by the M-protein doubling time, has been measured repeatedly on myeloma patients in unmaintained remissions (17). A progressive shortening in the M-protein doubling time was observed. Most of the patients who progressed to an M-protein doubling time of less than 30 days also developed marrow failure (5). This marrow failure apparently did not result from treatment, because the marrow remained cellular and the pancytopenia persisted after all therapy was stopped. About one third of myeloma patients progressed to develop marrow failure and died during the "acute terminal phase" of the disease (5). The M-protein doubling time during relapse correlates strongly with subsequent survival (17); short M-protein doubling times predict brief survival. The survival of myeloma patients appears to be determined largely by the time required for a progressive loss of myeloma growth control and marrow failure to occur.

A high incidence of acute leukemia has been reported in myeloma patients (17, 42-44). The NCI-C Clinical Trials group recorded six deaths from acute leukemia during three years of following 364 myeloma patients being treated with MP or MCBP. The expected number of acute leukemias in this group of patients, based on the 1974 incidence figures compiled by Statistics Canada, is 0.0489. The ratio of observed to expected acute leukemias is 122 (26). Two hypotheses have been proposed to explain the unusually high incidence of acute leukemia in myeloma patients: (a) Acute leukemia is induced by leukemogens, such as irradiation or alkylating agents, used in treatment; (b) Acute leukemia occurs as part of the natural history of myeloma. Conclusive evidence is not available to establish either of these hypotheses. I believe, however, that the available evidence favors the second.

To establish the first hypothesis it is necessary to demonstrate that the incidence of acute leukemia is significantly higher in myeloma patients treated with irradiation and alkylating agents than in a comparable group of untreated patients. Of course it would be unethical to do this study, because treatment with alkylating agents does prolong survival. And because of the prolonged survival of treated patients, it is not possible to do a retrospective analysis of the incidence of acute leukemia now as compared to what it was prior to the introduction of alkylating agents in the treatment of myeloma.

Favoring the second hypothesis is the occurrence of acute leukemia in untreated myeloma patients, and certain features occurring during the course of myeloma that resemble the progression of chronic granulocytic leukemia to the blast phase.

Acute leukemia has been reported to develop prior to the onset of therapy in at least 15 myeloma and 3 macroglobulinemia patients (43, 44). This suggests that the incidence of acute leukemia is increased in patients with plasma cell neoplasms, even in the absence of treatment.

The progressive increase in the myeloma growth rate observed during the course of the disease, which leads to marrow failure and acute leukemia, resembles the course of chronic granulocytic leukemia as it evolves to the acute blast phase. The acute leukemia that occurs in myeloma patients may possibly be part of the natural history of the disease, as it is in chronic granulocytic leukemia, polycythemia rubra vera, and myelofibrosis.

ACTIVE GROWTH CONTROL MECHANISMS IN PATIENTS WITH PLASMA CELL NEOPLASMS

Benign Monoclonal Immunoglobulinopathy

The plasma cell disease benign monoclonal immunoglobulinopathy (BMI) has been classified as a controlled proliferation of a single clone of immunoglobulin-producing cells (6). In this disorder, a clone of plasma cells (presumably derived from a single cell) expands until the population reaches a certain size (as many as 5×10^{11} cells), and then remains stable for many years. This is the commonest disorder responsible for a serum M-protein (45) and has been found in 1% of adults over the age of 25, 3% over 70, and 19% of nonagenarians (46). Progression of BMI to myeloma has been reported (47) but is uncommon. Hamburger & Salmon (48) were able to grow colonies of plasma cells from the marrow of 34 of 41 untreated or relapsing myeloma patients, 2 of 2 patients with macroglobulinemia, and 2 of 3 patients with BMI. Plasma cell colonies were not recovered from any of the marrow cultures of 10 patients without a plasma cell disorder, nor

from one patient with reactive plasmacytosis. Thus, in BMI there appears to be a population of cells capable of producing an M-protein and of forming colonies of plasma cells in vitro, as myeloma cells do. This population expands to a certain cell number and then remains stable for many years. The major difference from myeloma is that the growth of the monoclonal is controlled in BMI, so that a progressive increase in cell number and the development of plasmacytomas or osteolytic lesions does not occur.

Myeloma Remissions

In some ways the status of patients during treatment-induced remissions resembles that of patients with BMI. The M-protein level falls progressively during remission-induction to reach a plateau, and then remains stable throughout the remission. The duration of this stable phase is not prolonged by maintenance therapy. Some of these remissions persist for unusually prolonged periods. The remission that has lasted for more than eight years following 225 rads of total body irradiation (17) has been mentioned. In addition, I have observed unmaintained remissions of five years in a patient with IgG kappa myeloma, and in a patient producing only lambda light chains that has continued six years after melphalan was discontinued. Moreover, at least one long remission of more than five years has been reported following only one course of melphalan in a kappa light-chain disease patient (49). The persistence of an M-protein throughout these remissions, and the demonstration that 12 of 24 remission marrows contained myeloma stem cells capable of forming colonies of plasma cells (48), indicates that the myeloma clone persists but is controlled during remissions.

In both BMI and myeloma remissions there are large monoclonal populations of plasma cells (greater than 10^{10} cells) that remain stable for prolonged periods. Myeloma stem cells capable of forming colonies of plasma cells are present in both BMI and remission marrows. Unmaintained myeloma remissions last for variable periods, from a few months to many years, with a median duration of 11 months (41). A recurrence of progressive, uncontrolled growth of myeloma cells ends the remission in most myeloma patients. In contrast, progression from the stable monoclonal in BMI to myeloma is uncommon.

Suppressor T Cells

Paglieroni, MacKenzie & Caggiano (50) provided evidence that suggests that myeloma marrow contains a population of suppressor T cells. These investigators tested the reaction of autologous peripheral blood lymphocytes (PBL) against marrow cells in the one-way mixed leukocyte reaction. PBL did not respond to unfractionated myeloma marrow cells. There

was a strong reaction of PBL to purified myeloma cells, which could be suppressed by myeloma marrow lymphocytes. The removal of E-rosette-forming cells from the marrow lymphocyte fraction removed the suppressor cells, and return of the E-rosette-forming cells restored suppressor activity. This is the evidence for the presence of suppressor T cells in myeloma marrow.

Growth Control Mechanism Hypothesis

Further studies of the marrows of treated myeloma patients, and of BMI patients, suggest a possible growth control mechanism to explain the stable state of the monoclonal during myeloma remissions and in BMI (50). Suppressor cells were not found in the marrows of BMI patients, and appeared to be reduced in the marrows of myeloma patients following treatment. It is possible that the suppressor T cell activity demonstrated in the marrow of myeloma patients represents part of a growth control mechanism. The hypothesis can be formulated as follows. Progressive, uncontrolled growth of the monoclonal occurs only in the presence of marrow suppressor T cells (e.g. untreated or relapsing myeloma). It remains stable in patients with BMI, in whom suppressor T cells are not demonstrable in the marrow. Suppressor T cells appear to be reduced in the marrow of treated myeloma patients; possibly those who achieve prolonged, stable remissions are those who have the greatest reduction in suppressor T cells. Relapse occurs as suppressor T cells regenerate, and a progressive increase in myeloma growth rate occurs as the suppressor T cell population expands. A quantitative assay for suppressor T cells would facilitate the testing of this hypothesis.

SUMMARY

Studies during the past two decades have clarified the tumor biology of plasma cell neoplasms, identified the prognostic factors, and established a useful clinical staging system. The response rates and survival of myeloma patients treated with new agents and drug combinations are only minimally better than those achieved with melphalan and prednisone. The failure of more intensive treatment to improve survival, and the lack of a correlation between myeloma cell-kill and survival, suggests that the beneficial results of treatment may not be due to an effect of therapy on the myeloma cell mass.

Alternate hypotheses are required to explain the effect of treatment on plasma cell neoplasms. One alternate hypothesis holds that treatment reduces a population of suppressor T cells in myeloma patients and allows a growth control mechanism to be reestablished. This growth control mecha-

nism maintains the myeloma cell number stable until suppressor T cells regenerate sufficiently to allow the monoclonal to grow again. If research provides further support for a growth control mechanism in patients with plasma cell neoplasms, therapeutic research should be aimed at correcting the faulty control mechanisms, rather than at trying to reduce the tumor plasma cell number.

Literature Cited

1. Carbone, P. P., Kellerhouse, L. E., Gehan, E. A. 1967. Plasmacytic myeloma: A study of the relationship of survival to various clinical manifestations and anomalous protein type in 112 patients. *Am. J. Med.* 42:937-48
2. Medical Research Council's Working Party for Therapeutic Trials in Leukemia. 1973. Report of the first myelomatosis trial. Part I. Analysis of presenting features of prognostic significance. *Br. J. Haematol.* 24:123-39
3. Durie, B. G. M., Salmon, S. E. 1975. A clinical staging system for multiple myeloma. Correlation of measured myeloma cell mass with presenting clinical features, response to treatment and survival. *Cancer* 36:842-54
4. Alexanian, R., Balcerzak, S., Bonnet, J. D., Gehan, E. A., Haut, A., Hewlett, J. S., Monto, R. W. 1975. Prognostic factors in multiple myeloma. *Cancer* 36:1192-1201
5. Bergsagel, D. E., Pruzanski, W. 1975. Treatment of plasma cell myeloma with cytotoxic agents. *Arch. Intern. Med.* 135:172-76
6. Bergsagel, D. E., Pruzanski, W. 1978. Immunoglobulins in diagnosis and monitoring of neoplasia. In *Immunodiagnosis of Cancer*, ed. R. B. Herberman, K. R. McIntire. New York: Dekker. In press
7. Krajny, M., Pruzanski, W. 1976. Waldenström's macroglobulinemia: A review of 45 cases. *Can. Med. Assoc. J.* 114:899-905
8. Acute Leukemia Group B. 1975. Correlation of abnormal immunoglobulin with clinical features of myeloma. *Arch. Intern. Med.* 135:62-66
9. Alexanian, R., Haut, A., Khan, A. V., Lane, M., McKelvey, E. M., Migliore, P. J., Stuckey, W. J. Jr., Wilson, H. E. 1969. Treatment for multiple myeloma: Combination chemotherapy with different melphalan dose regimes. *J. Am. Med. Assoc.* 208:1680-85
10. Shustik, C., Bergsagel, D. E., Pruzanski, W. 1976. κ and λ light chain disease: Survival rates and clinical manifestations. *Blood* 48:41-51
11. Jancelewicz, Z., Takatsuki, K., Sugai, S., Pruzanski, W. 1975. IgD multiple myeloma: Review of 133 cases. *Arch. Intern. Med.* 135:87-93
12. Salmon, S. E., Smith, B. A. 1970. Immunoglobulin synthesis and total body tumor cell number in IgG myeloma. *J. Clin. Invest.* 49:1114-21
13. Blokhin, N., Larionov, L. F., Perevodchikova, N. I., Chebotareva, L., Merkulova, N. 1958. Clinical experiences with sarcosyl in neoplastic diseases. *Ann. NY Acad. Sci.* 68:1128-32
14. Bergsagel, D. E. 1962. Phase II trials of Mitomycin C, AB 100, NSC 1026, L-sarcosyl, and meta-sarcosyl, in the treatment of multiple myeloma. *Cancer Chemother. Rep.* 16:261-66
15. Bergsagel, D. E., Sprague, C. C., Austin, C., Griffith, K. M. 1962. Evaluation of new chemotherapeutic agents in the treatment of multiple myeloma. IV. L-phenylalanine mustard (NSC-8806). *Cancer Chemother. Rep.* 21:87-99
16. Salmon, S. E. 1973. Immunoglobulin synthesis and tumor kinetics of multiple myeloma. *Semin. Hematol.* 10:135-47
17. Bergsagel, D. E. 1977. Assessment of the response of mouse and human myeloma to chemotherapy and radiotherapy. In *Growth Kinetics and Biochemical Regulation of Normal and Malignant Cells* (Univ. Texas Syst. Cancer Cent. M.D. Anderson Hosp. Tumor Inst., 29th Ann. Symp. Fundam. Cancer Res., 1976), ed. B. Drewinko, R. M. Humphrey, pp. 705-17. Baltimore: Williams & Wilkins. 900 pp.
18. Abraham, D., Carbone, P. P., Vendetti, J. M., Kline, I., Goldin, A. 1967. An evaluation of chemical agents against the plasma cell tumor LPC-1 in mice. *Biochem. Pharmacol.* 16:665-73
19. Samuels, M. L., Leary, W. V., Alexanian, R., Howe, C. D., Frei, E. III. 1967. Clinical trials with *N*-isopropyl- α -(2-methylhydrazino)-*p*-toluamide hydrochloride in malignant lymphomas

- and other disseminated neoplasia. *Cancer* 20:1187-94
20. Moon, J. H., Edmonson, J. H. 1970. Procarbazine (NSC-77213) and multiple myeloma. *Cancer Chemother. Rep.* (Pt. 1) 54:245-48
 21. Salmon, S. E. 1975. Expansion of the growth fraction in multiple myeloma with alkylating agents. *Blood* 145: 119-29
 22. O'Bryan, R. M., Luce, J. K., Talley, R. W., Gottlieb, J. A., Baker, L. H., Bonadonna, G. 1973. Phase II evaluation of adriamycin in human neoplasia. *Cancer* 32:1-17
 23. Alberts, D. S., Salmon, S. E. 1975. Adriamycin in the treatment of alkylator-resistant multiple myeloma. *Cancer Chemother. Rep.* 59:345-50
 24. Ogawa, M., Gale, G. R., Meischan, S. J., Cooke, V. A. 1976. Effects of dinitrato (1,2-diaminocyclohexane) platinum (NSC 239851) on murine myeloma and hemopoietic precursor cells. *Cancer Res.* 36:3185-88
 25. Alexanian, R., Bonnet, J., Gehan, E., Haut, A., Hewlett, J., Lane, M., Monto, R., Wilson, H. 1972. Combination chemotherapy for multiple myeloma. *Cancer* 30:382-89
 26. Bergsagel, D. E., Bailey, A. J., Langley, G. R., MacDonald, N., White, D. F., Miller, A. B. 1977. Treatment of plasma cell myeloma with prednisone and sequential, alternating, or concurrent schedules of melphalan, cyclophosphamide and carmustine. *Proc. Am. Assoc. Cancer Res. Am. Soc. Clin. Oncol.* 18:329 (Abstr.)
 27. Alexanian, R., Salmon, S., Bonnet, J., Gehan, J., Haut, A., Weich, J. 1977. Combination therapy for multiple myeloma. *Cancer* 40:2765-71
 28. Southwest Oncology Group. 1975. Remission maintenance therapy for multiple myeloma. *Arch. Intern. Med.* 135:147-52
 29. Bergsagel, D. E. 1972. Plasma cell myeloma—An interpretive review. *Cancer* 30:1588-94
 30. Raisz, L. G., Luben, R. A., Mundy, G. R., Dietrich, J. W., Horton, J. E., Trummel, C. L. 1975. Effect of osteoclast activating factor from human leukocytes on bone metabolism. *J. Clin. Invest.* 56:408-13
 31. Bergsagel, D. E., Ogawa, M., Librach, S. L. 1975. Mouse myeloma. A model for studies of cell kinetics. *Arch. Intern. Med.* 135:109-13
 32. Bergsagel, D. E., Cowan, D. H., Hasselback, R. 1972. Plasma cell myeloma: Response of melphalan-resistant patients to high-dose intermittent cyclophosphamide. *Can. Med. Assoc. J.* 107:851-55
 33. Valeriote, F., Bruce, W. R., Meeker, B. E. 1968. Synergistic action of cyclophosphamide and 1,3-bis (2-chloroethyl)-1-nitrosourea on a transplanted murine lymphoma. *J. Natl. Cancer Inst.* 40:935-44
 34. Lin, H., Bruce, W. R. 1972. Chemotherapy of the transplanted KHT fibrosarcoma in mice. *Ser. Haematol.* 5:89-104
 35. Harley, J., McIntyre, O. R., Pajak, T. F. 1977. Improved survival of poor risk myeloma patients receiving combination alkylating agent therapy. *Blood* 50 (Suppl. 1):192 (Abstr.)
 36. Tattersall, M. H. N., Jarman, M., Newlands, E. S., Holyhead, L., Milstead, R. A. V., Weinburg, A. 1978. Pharmacokinetics of melphalan following oral or intravenous administration in patients with malignant disease. *Eur. J. Cancer* 14:507-13
 37. Alberts, D. S., Chang, S. Y., Chen, H.-S., Gross, J. F., Walson, P. D., Moon, T. E., Salmon, S. E. 1978. Variability of melphalan absorption in man. *Proc. Am. Assoc. Cancer Res. Am. Soc. Clin. Oncol.* 19:334 (Abstr.)
 38. Drewinko, B., Brown, B. W., Humphrey, R., Alexanian, R. 1974. Effect of chemotherapy on the labeling index of myeloma cells. *Cancer* 34:526-31
 39. Case, D. C. Jr., Lee, B. J. III., Clarkson, B. D. 1977. Improved survival times in multiple myeloma treated with melphalan, prednisone, cyclophosphamide, vincristine and BCNU: M-2 protocol. *Am. J. Med.* 63:897-903
 40. Bergsagel, D. E. 1971. Total body irradiation for myelomatosis. *Br. Med. J.* 2:325
 41. Alexanian, R., Gehan, E., Haut, A., Saiki, J., Weick, J. 1978. Unmaintained remissions in multiple myeloma. *Blood* 51:1005-11
 42. Khaleeli, M., Keane, W. M., Lee, G. R. 1973. Sideroblastic anemia in multiple myeloma: A preleukemic change. *Blood* 41:17-25
 43. Rosner, F., Grunwald, H. 1974. Multiple myeloma terminating in acute leukemia: Report of 12 cases and review of the literature. *Am. J. Med.* 57:927-39
 44. Cleary, B., Binder, R. A., Kales, A. N., Veltri, B. J. 1978. Simultaneous presentation of acute myelomonocytic leukemia and multiple myeloma. *Cancer* 41:1381-86

45. Axelsson, U., Bachmann, R., Hällén, J. 1966. Frequency of pathological proteins (M Components) in 6995 sera from an adult population. *Acta Med. Scand.* 179:235-47
46. Englišová, M., Engliš, M., Kyrál, V., Kouřilek, K., Dvořák, D. 1968. Changes of immunoglobulin synthesis in old people. *Exp. Gerontol.* 3:125-27
47. Waldenström, J. G. 1973. Benign monoclonal gammopathies. In *Multiple Myeloma and Related Disorders*, ed. H. A. Azur, M. Potter, 1:247-86. Hagerstown, Md: Harper & Row. 428 pp.
48. Hamburger, A., Salmon, S. 1977. Primary bioassay of human myeloma stem cells. *J. Clin. Invest.* 50:846-54
49. Von Schéele, C. 1976. Light chain myeloma with features of adult Fanconi syndrome: Six years remission following one course of melphalan. *Acta Med. Scand.* 199:533-37
50. Paglieroni, T., MacKenzie, M. R., Caggiano, V. 1978. Multiple myeloma: An immunologic profile. II. Bone marrow studies. *J. Natl. Cancer Inst.* 60:943-50

PATTERNS OF CIRCULATING ◆7334 COMPLEMENT IN RENAL DISEASES

Edmund J. Lewis, M.D.

Department of Medicine, Rush Medical College, Chicago, Illinois 60612

The complement system consists of a group of plasma proteins that participate in antibody-mediated cytolysis. The sequential activation of the components of the complement system results in the generation of biologically active products, which in turn mediate many aspects of the immune and inflammatory responses (1). The ability of circulating immune complexes to activate complement has resulted in the clinical use of complement component levels as indicators of immune activity. Serum complement levels deviate from the norm in many disease states in which the kidneys are affected. An inverse correlation exists between the serum levels of complement components and circulating levels of immune complexes in several renal diseases (2).

ACQUIRED ABNORMALITIES OF THE COMPLEMENT SYSTEM

Systemic Lupus Erythematosus

Many serum complement components show decreased levels in active systemic lupus erythematosus (3-5). As a general rule the serum levels of C4 and C3 correlate best with activity of the glomerular lesion (6-8). Depression of serum C4 may occur prior to the depression of other complement components. In addition, with the induction of a remission, C4 tends to return to normal more slowly than does C3. As systemic lupus erythematosus is taken to be the prototypical immune complex-mediated disease, the degree of hypocomplementemia has been interpreted to represent an indirect quantitative measure of circulating immune complex levels. The degree of depression of serum C3 correlates with the severity of glomerular disease (9). Patients with mesangial and mild focal glomerulonephritis show

less marked hypocomplementemia than do patients with evidence of glomerular necrosis or diffuse proliferative glomerulonephritis. Patients with membranous lupus glomerulonephritis may also have decreased serum C3 levels; however, normal values have also been observed in this group (9, 10). Depressed levels of C1q and the terminal complement components (C5–C9) (11, 12), as well as decreased levels of factor B and properdin also occur (13, 14). It therefore appears that activation of both classical and alternative complement pathways can be reflected in serum complement profiles during active lupus nephritis.

Hypocomplementemia in lupus is associated with elevation of the catabolic rate of C3 (12, 15–17). This is presumably the result of immune utilization. The synthetic rate of C3 has also been measured in patients with active and inactive lupus and is frequently below normal. However, while there does not appear to be an absolute correlation between disease activity and synthetic rate, diminished synthesis has been reported in patients with more severe disease activity. Catabolic products of C3 activation may possibly exert a negative feedback upon C3 synthesis (16). An elevated C3 synthetic rate has also been reported in some patients with active lupus (17). The serum C3 level at any given time therefore depends upon the balance of the catabolic and synthetic rates. This presumably explains normal C3 levels in a few patients with active lupus nephritis who have a markedly increased catabolic rate.

Further evidence of the degree of immune activity of lupus has been derived from the detection of breakdown products of C3 and factor B in the serum of patients with active systemic lupus erythematosus (13, 14). In these patients there is an inverse correlation between the serum C3 level and the quantity of circulating C3 breakdown products. Serial determinations of CH50, C3, and C4 are all of prognostic value in the clinical management of these patients (6–8, 18). A return to normocomplementemia was associated with stable or improved parameters of renal function (18), while persistent hypocomplementemia was associated with deterioration of renal function (18, 19).

Immune Complex Glomerulonephritis Associated with Infections

Circulating immune complexes and immune complex-mediated glomerulopathy reportedly occur in association with acute and subacute bacterial endocarditis (19–24), infected ventriculoatrial shunts (25, 26), visceral pyogenic suppuration, pneumococcal pneumonia (28), and with the therapeutic administration of heat-killed *Corynebacterium parvum* (29). Decreased serum C3 levels and the presence of circulating immune complexes accompany active infection in these disease states. Complement profiles also reveal depressed C1q and C4 levels (25). With antibiotic therapy and the eradica-

tion of bacteremia, the levels of C1q, C4, and C3 return to normal (22, 25, 27). The incidence of hypocomplementemia is greater in bacterial endocarditis than in other septic processes. Patients with subacute bacterial endocarditis who exhibit clinical evidence of glomerulonephritis, e.g. elevated serum creatinine, proteinuria, or hematuria, have a significantly higher incidence of hypocomplementemia than do endocarditis patients without evident renal involvement (22). The hypocomplementemia accompanying the bacteremias appears to be due to the activation of the complement system by the classical pathway (25). However, alternative pathway activation was also documented in a patient with pneumococcal sepsis and nephritis (28).

Essential Mixed Cryoglobulinemia

Essential mixed cryoglobulinemia is believed to be an immune complex-mediated disease entity (30). The cryoprecipitating immunoglobulins usually include the IgG and IgM classes, although IgA has also been reported in a small number of cases. One of these immunoglobulin classes often has rheumatoid factor activity: This suggests that the cryoglobulin represents an immune complex in which both antigen and antibody are immunoglobulin molecules. Other antigen-antibody systems have been reported in mixed cryoglobulinemic states, including DNA-anti-DNA, hepatitis B-anti-hepatitis B, and coccidioides antigen and antibody (31-33).

Immune complex deposition appears to explain clinical manifestations of this syndrome such as purpura, arthralgias, and proliferative glomerulonephritis. Decreased serum C3 and CH50 are frequently recorded (34-39). In some instances however, essential mixed cryoglobulinemia may be associated with depressed levels of C1q and C4, while normal levels of C3 and terminal components are present (39). This latter finding has not been precisely explained. There is no clear correlation between the degree of hypocomplementemia and either the clinical activity of the disease or the level of circulating cryoglobulins (39).

Membranoproliferative Glomerulonephritis

Membranoproliferative glomerulonephritis (MPGN) presents a characteristic histologic lesion (40). Two types have been defined on the basis of electron microscopic studies: Subendothelial deposits are noted in Type I; the second variant (Type II or dense deposit disease) is recognized by deposition of electron-dense material in the lamina densa of the glomerular basement membrane. The distinctive differences in these morphologic patterns suggest disparate pathogenetic mechanisms. Until recently most studies of serum complement components in MPGN did not take into account the variability of the histologic lesion. Patients with MPGN usually have hypocomplementemia at some time during the disease course (41). Initial

reports of low C3 terminal components with C1q, C4, and C2 remaining normal suggested alternative complement pathway activation (3–5, 43). Recent studies confirmed this finding in patients with dense deposit disease. However, patients with the Type I variant may also have low C4 levels (40). Therefore both activation pathways may be operative in these latter patients.

In addition to the distinctive complement profile, other evidence of alternative pathway activation may include decreased serum properdin and factor B levels (13, 14). Breakdown products of factor B and C3 were detected in the serum, which thus implies an increased *in vivo* turnover of these components (13, 14). The factor believed to be responsible for C3 activation in the serum of these patients is an IgG capable of activating the alternative complement pathway. It is designated C3 nephritic factor (C3NeF) (42) and is associated primarily with Type II MPGN. There is an inverse correlation between C3NeF activity and serum C3 (45). In patients with Type I MPGN, serum C3 and C4 levels tend to fluctuate and appear to vary independently of one another. Patients with Type II MPGN tend to have persistently depressed C3 levels. In neither Type I nor Type II MPGN do serum complement component levels correlate with disease activity in the glomerulus (40, 44–46).

Acute Poststreptococcal Glomerulonephritis

The morphologic analogy between the glomerular lesion of acute poststreptococcal glomerulonephritis (APSGN) and that of experimental acute serum sickness implies an immune complex pathogenesis for APSGN. Depressed complement levels are characteristically noted in these patients. To date there has been no convincing evidence to suggest that hypocomplementemia reflects the degree of immunopathologic activity in the kidney (47). Unlike the nephritis of systemic lupus erythematosus, complement consumption and hypocomplementemia have not been correlated with either circulating levels of immune complexes or with clinical activity of the glomerular lesion. While decreased C3 levels are a characteristic feature of the disease, normal complement levels may be found in 10–20% of cases (47–50). There is no apparent difference in the glomerular lesion or clinical course of patients with normocomplementemia as compared with hypocomplementemic patients. There is an initial depletion of the classical pathway components; however, later in the course of the disease C3 depression may be an isolated finding (3, 4, 51, 52). C1q and C4 may be significantly reduced during the first days of the disease, but the serum C3 level usually remains low for 4–6 weeks. Properdin levels may also be decreased. Clinical improvement of the renal lesion usually precedes the return to normocomplementemia, though total resolution of the urinary findings is usually not

complete when the C3 returns to normal (50, 51). While there is no correlation between the severity of the glomerulonephritis and the degree of C3 depression, several instances of unresolved hematuria and proteinuria with prolonged hypocomplementemia have been recorded (5, 50). Breakdown products of C3 may be detected in the circulation early in the course of the disease (14, 53).

Thrombotic Glomerulopathies

Low serum levels of C3 have been inconsistently noted in disease states associated with disseminated intravascular coagulation (DIC) (54, 55). Several mechanisms of complement activation could explain hypocomplementemia in these patients. Antigen-antibody complexes, endotoxin, bacterial lipopolysaccharide, and red blood cell stroma all represent factors capable of activating C3 by the alternative pathway during DIC. Reduced levels of C1q, C4, C3, and factor B have been reported in a few patients with the hemolytic-uremic syndrome (HUS) (56–58). When HUS is associated with a bacterial infection, as can occur during an outbreak of shigellosis, depressed levels of C3 and C4 appear to correlate with the presence of circulating immune complexes. The latter instance may represent classical pathway activation (59).

Idiopathic Nephrotic Syndrome

Altered metabolism of complement components accompanies the idiopathic nephrotic syndrome (INS) (minimal change glomerulopathy). It is usually explained by mechanisms other than immune utilization. The early classical pathway serum complement components are generally present in normal concentrations, with the exception of occasionally decreased levels of C1 or C1q (3, 5). Significantly decreased concentrations of C8 (11), factor B (60, 61), and C3b inactivator (C3bINA) (61) are associated with exacerbations of INS. In one study, factor B levels correlated with the serum albumin, and it was suggested that depression of this component was due to urinary loss (60). However, this correlation has not been confirmed (61). Total hemolytic complement levels may be low, but C3 is normal in these patients (3). Despite the recorded variations in complement component levels, the complement system is not clearly involved in the pathogenesis of the glomerular lesion of INS.

HERITABLE ABNORMALITIES OF THE COMPLEMENT SYSTEM

Genetically determined deficiency states have been described for a number of individual complement components (62). As the component deficiency

is usually an isolated absence of a single factor, the recognition of the abnormality usually rests on the finding of a very low or absent level of CH50. The clinical syndromes associated with genetic defects of individual complement components in man vary according to the level at which complement activation is interrupted.

Deficiency of Early Classical Pathway Components

Syndromes similar or identical to classical systemic lupus erythematosus have been described in a number of patients with isolated deficiency of an early classical pathway component. In one instance of C1r deficiency, glomerulonephritis led to end-stage renal disease (63). In a second, clinical evidence of glomerulonephritis was minimal; however, renal biopsy revealed immune complex deposition in the glomerular mesangium (64). The lupus syndrome associated with homozygous C4 deficiency is associated with glomerulonephritis and the nephrotic syndrome (65). Renal biopsy revealed changes compatible with lupus glomerulonephritis, including mesangial and subendothelial electron-dense deposits. In addition to the absence of serum C4, C3 and factor B are diminished in association with active nephritis.

The most common of these rare deficiency states, and the best studied, is that of the homozygous C2 deficient patients. Lupus syndromes frequently accompany this defect. Although the nephritis is usually mild in these patients, severe glomerulonephritis has been reported (66–68). While the majority of these patients appear to have limited inflammatory manifestations associated with the interruption of the classical activation pathway, effective inflammatory damage can apparently be mediated by the intact alternative pathway (67, 68). Hence, serum levels of properdin, factor B, C3, and the terminal components may be diminished in association with genetic absence of an early classical pathway component.

Deficiency of C1 Inhibitor

In addition to the classic pattern of hereditary angioedema (HAE) associated with C1 inhibitor deficiency, several patients also exhibited a lupus syndrome (62). Diffuse proliferative glomerulonephritis was reported in one of these patients (69).

C3 Deficiency States

Patients with inherited absence of C3 frequently have life-threatening bacterial infections (62). Renal disease has not been reported in these patients. However, acquired chronic depression of C3 has been well described in patients who develop partial lipodystrophy (70–73). There is a high rate of

development of membranoproliferative glomerulonephritis of the "dense deposit disease" type in these patients. C3NeF is present in the serum and depressed C3 is reportedly a family trait (74). The relationship of hypocomplementemia to the predisposition to partial lipodystrophy, and the susceptibility of these patients to membranoproliferative glomerulonephritis, remains unexplained. The chronic complement deficiency may predispose to an infectious process that mediates the glomerular lesion (71). However, the morphology of dense deposit disease is not analogous to recognized models of immune complex-mediated glomerulopathy. Hypocomplementemia in these patients is characterized by depressed C3 with normal C1, C4, and C2 levels.

Deficiencies of the Terminal Components

Individual defects of C5–C8 have been described and, in general, these individuals are normal. An increased susceptibility to infection by gram-negative bacteria, such as *neisseria* has been suggested. Chronic renal failure was reported in one patient with isolated absence of C7 (75). The clinical characteristics of this patient did not suggest a primary glomerular lesion and were compatible with chronic interstitial nephritis. A patient with absence of C8 was reported with a systemic lupus erythematosus syndrome and a histologic picture compatible with membranous glomerulopathy (76). This patient had a 20-year history of nephrotic syndrome and slowly progressing renal failure. During the latter part of this patient's course, she had normal serum C3 and predictably absent CH50.

CONCLUSION

The determination of serum complement components has elucidated the immunopathogenetic mechanisms involved in several renal diseases. The correlation between immune complex levels, disease activity, and complement component depletion in such diseases as systemic lupus erythematosus and the bacteremias associated with bacterial endocarditis and infected ventriculoatrial shunts make these serologic parameters highly relevant in the management of these patients. Complement component levels in other diseases such as poststreptococcal glomerulonephritis and membranoproliferative glomerulonephritis have provided considerable insight into the relevance of the complement system to disease mechanisms, particularly with respect to the involvement of the alternative pathway. The apparent predisposition of patients with inherited deficiencies of isolated complement components to lupus-like syndromes and glomerulonephritis remains a fascinating area for the development of new insights into the biological functions of the complement system.

Literature Cited

1. Müller-Eberhard, H. J. 1975. Complement. *Ann. Rev. Biochem.* 44:697-724
2. Woodroffe, A. J., Border, W. A., Theofilopoulos, A. N., Gotze, O., Glasscock, R. J., Dixon, F. J., Wilson, C. B. 1977. Detection of circulating immune complexes in patients with glomerulonephritis. *Kidney Int.* 12: 268-78
3. Gewurz, H., Pickering, R. J., Mergenhagen, S. E., Good, R. A. 1968. The complement profile in acute glomerulonephritis, systemic lupus erythematosus and hypocomplementemic chronic glomerulonephritis. *Int. Arch. Allergy Appl. Immunol.* 34:556-70
4. Kohler, P. F., Ten Bensel, R. 1969. Serial complement component alterations in acute glomerulonephritis and systemic lupus erythematosus. *Clin. Exp. Immunol.* 4:191-202
5. Lewis, E. J., Carpenter, C. B., Schur, P. H. 1971. Serum complement component levels in human glomerulonephritis. *Ann. Intern. Med.* 75:555-60
6. Cameron, J. S., Vick, R. M., Ogg, C. S., Seymour, W. M., Chantler, C., Turner, D. R. 1973. Plasma C3 and C4 concentrations in management of glomerulonephritis. *Br. Med. J.* 2:668-72
7. Williams, D. G., Peters, D. K., Fallows, J., Petrie, A., Kourilsky, O., Morel-Maroger, L., Cameron, J. S. 1974. Studies of serum complement in the hypocomplementaemic nephritides. *Clin. Exp. Immunol.* 18:391-405
8. Cameron, J. S., Lessof, M. H., Ogg, C. S., Williams, B. D., Williams, D. G. 1976. Disease activity in the nephritis of systemic lupus erythematosus in relation to serum complement concentrations. *Clin. Exp. Immunol.* 25:418-27
9. Hill, G. S., Hinglais, N., Tron, F., Bach, J.-F. 1978. Systemic lupus erythematosus. Morphologic correlations with immunologic and clinical data at the time of biopsy. *Am. J. Med.* 64:61-86
10. Donadio, J. V. Jr., Holley, K. E. 1977. Membranous lupus nephropathy: A clinicopathological study. *Medicine (Baltimore)* 56:527-36
11. Geiger, H., Good, R. A., Day, N. K. 1975. A study of complement components C3, C5, C6, C7, C8 and C9 in chronic membranoproliferative glomerulonephritis, systemic lupus erythematosus, poststreptococcal nephritis, idiopathic nephrotic syndrome and anaphylactoid purpura. *Z. Kinderheilkd.* 119:269-78
12. Hunsicker, L. G., Ruddy, S., Carpenter, C. B., Schur, P. H., Merrill, J. P., Müller-Eberhard, H. J., Austen, K. F. 1972. Metabolism of third complement component (C3) in nephritis. Involvement of the classic and alternate (properdin) pathways for complement activation. *N. Engl. J. Med.* 287:835-40
13. McLean, R. H., Michael, A. F. 1973. Properdin and C3 proactivator: alternate pathway component in human glomerulonephritis. *J. Clin. Invest.* 52: 634-44
14. Perrin, L. H., Lambert, P. H., Miescher, P. A. 1975. Complement breakdown products in plasma from patients with systemic lupus erythematosus and patients with membranoproliferative or other glomerulonephritis. *J. Clin. Invest.* 56:165-76
15. Ruddy, S., Carpenter, C. B., Chin, K. W., Knotsman, J. N., Soter, N. A., Gotze, O., Müller-Eberhard, H. J., Austen, K. F. 1975. Human complement metabolism: An analysis of 144 studies. *Medicine (Baltimore)* 54:165-78
16. Charlesworth, J. A., Williams, D. G., Sherington, E., Lachmann, P. J., Peters, D. K. 1974. Metabolic studies of the third component of complement and the glycine-rich beta glycoprotein in patients with hypocomplementemia. *J. Clin. Invest.* 53:1578-87
17. Petz, L. D., Powers, R., Fries, J., Cooper, N. R., Homan, H. R. 1977. The in vivo metabolism of the third component of complement in systemic lupus erythematosus. *Arthritis Rheum.* 20: 1304-13
18. Appel, A. E., Sablay, L. B., Golden, R. A., Barland, P., Grayzel, A. I., Bank, N. 1978. The effect of normalization of serum complement and anti-DNA antibody on the course of lupus nephritis. *Am. J. Med.* 64:274-83
19. Hecht, B., Siegel, N., Adler, M., Kashgarian, M., Hayslett, J. P. 1976. Prognostic indices in lupus nephritis. *Medicine (Baltimore)* 55:163-81
20. Gutman, R. A., Striker, G. E., Gilliland, B. C., Cutler, R. E. 1972. The immune complex glomerulonephritis of bacterial endocarditis. *Medicine (Baltimore)* 51:1-25
21. Beaufrils, M., Gibert, C., Morel-Maroger, L., Sraer, J. D., Kanfer, A., Meyrier, A., Kourilsky, O., Vachon, F., Richet, G. 1977. Glomerulonephritis in severe bacterial infections with and without endocarditis. *Adv. Nephrol.* 7:217-34

22. Pelletier, L. L. Jr., Petersdorf, R. G. 1977. Infective endocarditis: A review of 125 cases from the University of Washington Hospitals, 1963-1972. *Medicine (Baltimore)* 56:287-314
23. Levy, R. L., Hong, R. 1973. The immune nature of subacute bacterial endocarditis (SBE) nephritis. *Am. J. Med.* 54:645-52
24. Kim, Y., Michael, A. F. 1978. Chronic bacteremia and nephritis. *Ann. Rev. Med.* 29:319-25
25. Strife, C. F., McDonald, B. M., Ruley, E. J., McAdams, A. J., West, C. D. 1976. Shunt nephritis: The nature of the serum cryoglobulins and their relation to the complement profile. *J. Pediatr.* 88:403-13
26. Kaufman, D. B., McIntosh, R. 1971. The pathogenesis of the renal lesion in a patient with streptococcal disease, infected ventriculoatrial shunt, cryoglobulinemia and nephritis. *Am. J. Med.* 50:262-68
27. Beaufils, M., Morel-Maroger, L., Sraer, J.-D., Kanfer, A., Kourilsky, O., Richet, G. 1976. Acute renal failure of glomerular origin during visceral abscesses. *N. Engl. Med. J.* 295:185-89
28. Hyman, L. R., Jems, E. H., Hill, G. S., Zimmerman, S. W., Burkholder, P. M. 1975. Alternate C3 pathway activation in pneumococcal glomerulonephritis. *Am. J. Med.* 58:810-14
29. Dosik, G. M., Gutterman, J. U., Hersh, E. M., Akhtar, M., Sonoda, T., Horn, R. G. 1978. Nephrotoxicity from cancer immunotherapy. *Ann. Intern. Med.* 89:41-46
30. Brouet, J.-C., Clavel, J.-P., Danon, F., Klein, M., Seligmann, M. 1974. Biologic and clinical significance of cryoglobulins. A report of 86 cases. *Am. J. Med.* 57:775-88
31. Roberts, J. L., Lewis, E. J. 1979. Identification of anti-native DNA antibodies in cryoglobulinemic states. *Am. J. Med.* 65:437-45
32. Levo, Y., Gorevic, P. D., Kassab, H. J. 1977. Association between hepatitis B virus and essential mixed cryoglobulinemia. *N. Engl. J. Med.* 296:1501-4
33. Gamble, C. N., Ruggles, S. W. 1978. The immunopathogenesis of glomerulonephritis associated with mixed cryoglobulinemia. *N. Engl. J. Med.* 299:81-84
34. Meltzer, M., Franklin, E. C., Elias, K., McCluskey, R. T., Cooper, N. 1966. Cryoglobulinemia. A clinical and laboratory study II. Cryoglobulins with rheumatoid factor activity. *Am. J. Med.* 40:837-56
35. Müller, S., Rother, V., Westerhausen, M. 1976. Complement activation by cryoglobulin. Evidence for a pathogenic role of an IgG (kappa) cryoglobulin in cutaneous vasculitis. *Clin. Exp. Immunol.* 23:233-41
36. Beaumont, J. L., Antonucci, M., Lagrue, G., Guedon, J., Perol, R. 1974. Nephrotic syndrome, monoclonal gammopathy and auto-immune hyperlipidemia. *Clin. Exp. Immunol.* 18:225-36
37. Cordonnier, D., Martin, H., Gros Lambert, P., Micouin, C., Chenais, F., Stoeber, P. 1975. Mixed IgG-IgM cryoglobulinemia with glomerulonephritis. *Am. J. Med.* 59:867-72
38. Bengtsson, V., Larsson, O., Lindstedt, G., Svalander, C. 1975. Monoclonal IgG cryoglobulinemia with secondary development of glomerulonephritis and nephrotic syndrome. *Q. J. Med.* 44:491-503
39. Tarantino, A., Anelli, A., Costantino, A., DeVecchi, A., Monti, G., Massaro, L. 1978. Serum complement pattern in essential mixed cryoglobulinaemia. *Clin. Exp. Immunol.* 32:77-85
40. Habib, R., Kleinknecht, C., Gubler, M.-C., Levy, M. 1973. Idiopathic membranoproliferative glomerulonephritis in children. Report of 105 cases. *Clin. Nephrol.* 1:194-214
41. West, C. D. 1973. Membranoproliferative hypocomplementemic glomerulonephritis. *Nephron* 11:134-46
42. Scott, D. M., Amos, N., Sissons, J. G. P., Lachmann, P. J., Peters, D. K. 1978. The immunoglobulin nature of nephritic factor (NeF) *Clin. Exp. Immunol.* 32:12-24
43. Peters, D. K., Martin, A., Weinstein, A., Cameron, J. S., Barratt, T. M., Ogg, C. S., Lachmann, P. J. 1972. Complement studies in membranoproliferative glomerulonephritis. *Clin. Exp. Immunol.* 11:311-20
44. Northway, J. D., McAdams, A. J., Forristal, J., West, C. D. 1969. A "silent" phase of hypocomplementemic persistent nephritis detectable by reduced serum β 1C-globulin levels. *J. Pediatr.* 74:28-38
45. Vallota, E. H., Forristal, J., Davis, N. C., West, C. D. 1972. The C3 nephritic factor and membranoproliferative nephritis: Correlation of serum levels of the nephritic factor with C3 levels, with therapy and with progression of the disease. *J. Pediatr.* 80:947-59

46. Vallota, E. H., Forristal, J., Spitzer, R. E., Davis, N. C., West, C. D. 1971. Continuing C3 breakdown after bilateral nephrectomy in patients with membranoproliferative glomerulonephritis. *J. Clin. Invest.* 50:552-58
47. Strife, C. F., McAdams, A. J., McEnery, P. T., Bove, K. E., West, C. D. 1974. Hypocomplementemic and normocomplementemic acute nephritis in children: A comparison with respect to etiology, clinical manifestations and glomerular morphology. *J. Pediatr.* 84:29-38
48. West, C. D., Northway, J. D., Davis, N. C. 1964. Serum levels of β 1C-globulin, a complement component, in the nephritides, lipoid nephrosis, and other conditions. *J. Clin. Invest.* 43:1507-17
49. Tina, L. V., D'Albora, J. B., Antonovych, T. T., Bellanti, J. A., Calcagno, P. L. 1968. Acute glomerulonephritis associated with normal serum β 1C-globulin. *Am. J. Dis. Child.* 115:29-36
50. Gotoff, S. P., Isaacs, E. W., Muehrcke, R. C., Smith, R. D. 1969. Serum β 1C-globulin in glomerulonephritis and systemic lupus erythematosus. *Ann. Intern. Med.* 71:327-33
51. Humair, L. M. 1968. β 1C-globulin and complement in nephritis. *Helv. Med. Acta* 34:279-96
52. McClean, R. H., Schragar, M. A., Rothfield, N. F., Berman, M. A. 1977. Normal complement in early poststreptococcal glomerulonephritis. *Br. Med. J.* 1:1326
53. Soothill, J. F. 1967. Altered complement component C3A (β 1C- β 1A) in patients with glomerulonephritis. *Clin. Exp. Immunol.* 2:83-92
54. Tomar, R. H., Kolchino, D. 1972. Complement and coagulation. Serum β 1C- β 1A in disseminated intravascular coagulation. *Thromb. Diath. Haemorrh.* 27:389-95
55. Bokisch, V. A., Top, F. H. Jr., Russell, P. K., Dixon, F. J., Müller-Eberhard, H. J. 1973. The potential pathogenic role of complement in dengue hemorrhagic shock syndrome. *N. Engl. J. Med.* 289:996-1000
56. Katz, J., Lurie, A., Kaplan, B. S., Krawitz, S., Metz, J. 1971. Coagulation findings in the hemolytic-uremic syndrome of infancy: Similarity to hyperacute renal allograft rejection. *J. Pediatr.* 78:426-34
57. Drukker, A., Winterborn, M., Bennett, B., Churg, J., Spitzer, A., Greifer, I. 1975. Recurrent hemolytic-uremic syndrome: A case report. *Clin. Nephrol.* 4:68-72
58. Stuhlinger, W., Kourilsky, O., Verroust, P. J., Sraer, J. D., Kanfer, A., Morel-Maroger, L. 1975. Immunological studies in the hemolytic-uremic syndrome of adults. *Proc. VIIth Int. Congr. Nephrol. Firenze, Italy, 1975.* Abstr. 410
59. Koster, F., Levin, J., Walker, L., Tung, K. S. K., Gilman, R. H., Rahaman, M., Majid, M. A., Islam, S., Williams, R. C. Jr. 1978. Hemolytic-uremic syndrome after Shigellosis. Relation to endotoxemia and circulating immune complexes. *N. Engl. J. Med.* 298:927-33
60. McLean, R. H. 1977. Decreased serum factor B concentration associated with decreased opsonization of *Escherichia coli* in the idiopathic nephrotic syndrome. *Pediatr. Res.* 11:910-16
61. Forristal, J., Iitaka, K., Vallota, E. H., West, C. D. 1977. Correlations between serum factor b and C3b inactivator levels in normal subjects and in patients with infections, nephrosis and hypocomplementemic glomerulonephritis. *Clin. Exp. Immunol.* 28:61-71
62. Agnello, V. 1978. Complement deficiency states. *Medicine (Baltimore)* 57:1-23
63. Pickering, R. J., Michael, A. F. Jr., Herdman, R. C., Good, R. A., Gewurz, H. 1971. The complement system in chronic glomerulonephritis: Three newly associated aberrations. *J. Pediatr.* 78:30-43
64. Moncada, B., Day, N. K. B., Good, R. A., Windhorst, D. B. 1972. Lupus-erythematosus-like syndrome with a familial defect of complement. *N. Engl. J. Med.* 286:689-93
65. Schaller, J. G., Gilliland, B. G., Ochs, H. D., Leddy, J. P., Agadoa, L. C., Rosenfeld, S. I. 1977. Severe systemic lupus erythematosus with nephritis in a boy with deficiency of the fourth component of complement. *Arthritis Rheum.* 20:1519-25
66. Rynes, R. I., Urizar, R. E., Rickering, R. J. 1978. Heterogeneity of the clinical syndrome in patients with systemic lupus erythematosus and genetic deficiency of the second component of complement. *Clin. Exp. Immunol.* 32:59-68
67. Roberts, J. L., Schwartz, M., Lewis, E. J. 1978. Hereditary C2 deficiency and systemic lupus erythematosus associated with severe glomerulonephritis. *Clin. Exp. Immunol.* 31:328-38
68. Gewurz, A., Lint, T. F., Roberts, J. L., Zeitz, H., Gewurz, H. 1978. Homozy-

- gous C2 deficiency with fulminant lupus erythematosus. Severe nephritis via the alternative complement pathway. *Arthritis Rheum.* 21:28-36
69. Rosenfeld, G. B., Partridge, R. E. H., Bartholemew, W., Murphy, W. H., Singleton, C. M. 1974. Hereditary angioneurotic edema (HANE) and systemic lupus erythematosus (SLE) in one of identical twin girls. *J. Allergy Clin. Immunol.* 53:68 (Abstr.)
70. Peters, D. K., Charlesworth, J. A., Sissons, J. G. P., Williams, D. G., Boulton-Jones, J. M., Evans, D. J. 1973. Mesangiocapillary nephritis partial lipodystrophy and hypocomplementaemia. *Lancet* 2:535-38
71. Peters, D. K., Williams, D. G. 1974. Complement and mesangiocapillary glomerulonephritis. Role of complement deficiency in the pathogenesis of nephritis. *Nephron* 13:189-91
72. Thompson, R. A., White, R. H. R. 1973. Partial lipodystrophy and hypocomplementemic glomerulonephritis. *Lancet* 2:679
73. Ipp, M. M., Minta, J. O., Gelfand, E. W. 1977. Disorders of the complement system in lipodystrophy. *Clin. Immunol. Immunopathol.* 7:281-87
74. Wakefield, M. A. 1976. The interrelationship between hypocomplementaemia, partial lipodystrophy and mesangiocapillary glomerulonephritis. *Postgrad. Med. J.* 52:651-52
75. Nemerow, G. R., Gewurz, H., Osofsky, S. G., Lint, T. 1968. Inherited deficiency of the seventh components of complement associated with nephritis. *J. Clin. Invest.* 61:1602-10
76. Jasin, E. 1977. Absence of the eighth component of complement in association with systemic-lupus erythematosus-like disease. *J. Clin. Invest.* 60:709-15

IMMUNIZATION AGAINST TYPHOID FEVER

◆7335

*John W. Warren, M.D. and Richard B. Hornick, M.D.*¹

Division of Infectious Diseases, Department of Medicine,
University of Maryland School of Medicine, Baltimore, Maryland 21201

Typhoid fever, an acute febrile illness caused by ingestion of and intestinal invasion by *Salmonella typhi*, is prevalent in areas of the world that lack sanitary water and sewage systems. Furthermore, educational levels and health knowledge in these countries are often insufficient to encourage measures that might prevent the fecal-oral transmission of the organism. The importance of sanitation and education in diminishing the incidence of typhoid fever and deaths due to the disease is suggested by epidemiological reports in the United States during the early part of this century. As public and private plumbing systems were introduced and popularized, the reported incidence of typhoid fever declined. Typhoid vaccination was never introduced on a wide scale in this country and thus had little bearing upon the incidence of the disease. Similarly, the introduction of chloramphenicol in 1948, while effective in decreasing mortality, was not associated with any discernible acceleration of the decline in reported cases (1).

The prospects for adequate health practices and sanitation in all areas of the world, at least in the reasonably near future, are bleak. Consequently, because prevention of exposure is a difficult immediate goal, prevention of disease by immunization becomes an acceptable alternative. The isolation of chloramphenicol-resistant (2) and ampicillin-resistant (3) strains in recent epidemics further emphasizes the need for such a measure. The purpose of this paper is to discuss the present status of immunization against typhoid fever. Recent information about immunology, particularly in the gastrointestinal tract, offers promising new pathways to explore for immu-

¹The authors gratefully acknowledge the expert reviews of this manuscript by Merrill J. Snyder and William E. Woodward.

nization purposes and thus is discussed first. Following this, we outline the known specific immunity against *Salmonella typhi*, the status of the parenteral vaccines, and the intriguing prospects for an oral vaccine.

IMMUNOLOGY OF TYPHOID FEVER

The explosion of knowledge in immunology over the last few decades has revealed the complex ties and interrelationships of the immune response in man. Humoral immunity is mediated by the B lymphocyte, which differentiates into a plasma cell and produces immunoglobulins specific for a particular antigen. Secretory IgG is the predominant class of antibody in serum and functions in a variety of ways including agglutination of organisms, opsonization, complement activation, and neutralization of toxins. IgA, a dimeric immunoglobulin possessing two sets of light chains and heavy chains, is the major immunoglobulin present in secretions, including those of the gastrointestinal tract. Though it may perform some of the same functions as IgG, its action in preventing bacterial adherence to epithelial cells is perhaps more interesting. IgM, a pentamer produced in large quantities after initial contact with an antigen, is also found in increased quantities in secretions in patients with selected IgA deficiency and thus may act as a "back-up" secretory antibody in the absence of IgA.

Cellular-mediated immunity, the major defense against viral and fungal infections and closely related to delayed hypersensitivity, is governed by the T lymphocyte with the help of the macrophage as both antigen processor and subsequent phagocyte. The complement cascade, activated by the classical or by the alternate pathway, provides an integral component of the immune response that results in chemotaxis, opsonization by phagocytes and, in some instances, bacteriolysis. Phagocytes, both macrophages and polymorphonuclear leukocytes, attracted by complement and other factors, move to the site of infection, recognize the invading agent with the help of opsonic antibodies and complement, and then ingest and kill the organism.

These mechanisms have been investigated in regard to enteric infections including typhoid fever. These investigations and those of nonspecific gastrointestinal mechanisms of defense are examined here particularly with regard to their importance in vaccine development.

Serum Antibodies

In 1896, Widal added serum from typhoid fever patients in serial dilution to broth cultures of *S. typhi* and noted gross flocculation at the higher concentrations (4). Since that demonstration and until recently, agglutinating antibodies in serum have been considered measures of immunity against the disease. Antibodies to three distinct types of salmonella antigens have been well studied. The O antigen is composed of a polysaccharide compo-

nent of the cell wall and is part of the salmonella endotoxin. The H antigen is a protein found in the flagella of the organism and, with the O antigen, is used to differentiate the salmonellas into more than 1400 serotypes. The Vi antigen, named for its association with virulence, is a carbohydrate in the envelope and inhibits agglutination of the organisms by anti-O antibodies. Although these antigens have been investigated for more than half a century, recent work suggests that serum antibodies against them are not the sole mechanism for specific immunity to typhoid fever (5).

Antibody response to these antigens has been studied in vaccine trials in endemic areas of the world as well as in American volunteers. Field trials conducted in Yugoslavia (6, 7), Guyana (8, 9), and Poland (10) demonstrated no consistent difference in O and Vi antibody response between persons receiving the more effective and those receiving the less effective vaccines. Volunteer trials similarly showed no relationship between immunity to *S. typhi* and antibody titers against O or Vi antigen (11-13). Both field and volunteer trials demonstrated, as a matter of fact, Vi antibody response in very few vaccinated individuals. The field trial in Poland tested a lipopolysaccharide vaccine that elicited only O antibody; this vaccine afforded no protection (14). The conclusion that must be entertained is that serum antibodies to O antigen and to Vi antigen are of little importance in reflecting immunity to typhoid fever.

However, several studies have provided clues that serum antibody to H antigen, the flagellar antigen, may be associated with subsequent resistance to infection. The field trials mentioned above each demonstrated that the more effective vaccine elicited significantly higher titers to the H antigen than did the less effective vaccines (5). In the American volunteers, the presence of preexisting serum antibodies to H antigen in the nonvaccinated controls was associated with a significantly lower attack rate than that found in controls with no anti-H antibody (15, 16). A third trial hinted at the possible importance of the H antigen: A parenteral vaccine that did not contain H antigen gave poor results in Egyptian school children (17). However, studies of chimpanzees demonstrated poor immunity following vaccination with purified H antigen (18). The comparability of typhoid fever in chimpanzees, usually self-limited enteritis without enteric fever, to that in man is questionable; consequently, perhaps also questionable is the relevance of studies of vaccine-induced immunity in chimpanzees to that in humans.

The possibility that serum antibodies to H antigen are associated with protection against typhoid fever is an intriguing one. Antibodies to the flagellum, the organelle of locomotion, may conceivably immobilize the organism, thus preventing access to the gut wall for subsequent penetration and invasion. Evidence for this thesis is available in *Vibrio cholerae* in that nonmotile mutants are less virulent because of diminished capacity to ad-

sorb to intestinal mucosa (19). Whether this disability is because of immobility itself or because of associated defects in adsorption is unknown, however.

Intestinal Antibodies

Central to the theory of immobilization or of other specific gastrointestinal immunity is the presence of antibodies in the gut lumen. Measurements of serum antibodies to H antigen may show a positive association with immunity but the actual protection may be taking place in the gut, not through serum IgG antibodies but through secretory IgA. The mucosa of the gastrointestinal tract forms a tremendous surface area wrapped around huge numbers of microorganisms. Antibody directed at specific pathogens is likely of fundamental importance in preventing invasion and infection (20–22). IgA, discovered in human serum in 1959, is the main immunoglobulin in human secretions such as saliva, tears, milk, bronchial secretions, urine, and bile (21). In intestinal secretions, IgA may be present in concentrations averaging 30 mg/100 ml of fluid (23). The source of this IgA appears to be lymphocytes that are found in the lamina propria just below the epithelial surface, as many as 400,000/mm³ of lamina propria, numbers far greater than all other Ig-containing cells (24).

The predominance of IgA and IgA-producing cells in the intestine has been assumed to represent an immune system against pathogens present in the gut lumen (25). Recent work has implied that an important immune function of IgA is to prevent adherence of organisms to a mucosal surface (26). The most reasonable primary event for invasion of a mucosal surface by an organism is contact with the mucosal cells. If this contact is prevented, then invasion can be prevented. After IgA is transported from the plasma cells in the lamina propria to the gut lumen, secretory IgA remains in the mucous coat lining the mucosal surface, an arrangement prompting the phrase “antiseptic paint” (27). By attaching to specific antigens on the surface of an organism, IgA may block those sites that would bind to epithelial cells. The organism cannot adhere to the epithelial surface and may be more easily disposed of by mechanical factors and other specific and nonspecific defense mechanisms. This is an attractive hypothesis for protection against intestinal infections and has been demonstrated for *V. cholerae* and *Escherichia coli* in the small intestine (28, 29). Such a process in typhoid fever is assumed but awaits proof.

Cell-Mediated Immunity

The role of cell-mediated immunity (CMI) in enteric diseases is even less well understood than that of antibodies. Studies demonstrated lymphocytes committed to CMI in Peyer's patches that can respond to antigens present

in the lumen (30). Intestinal biopsies in volunteers demonstrated the presence of lymphocytes and macrophages in the lamina propria (31). In humans recovery from typhoid fever is correlated with the development of CMI, as tested by leukocyte migration (32). CMI undoubtedly plays a role, and probably an important one, in the development of immunity to typhoid fever, but further work is necessary to measure its importance.

Phagocytosis

Phagocytosis is important in the immunological defense against infection by *S. typhi*. As noted, macrophages are present in the lamina propria of patients with typhoid (31). Phagocytic activity in the reticuloendothelial system in general is increased in mouse typhoid (33) and in typhoid fever in human volunteers (34). In vitro, the rate of phagocytosis of *S. typhi* by macrophages is increased by the addition of antiserum (35). However, phagocytosis may not always be associated with killing of typhoid bacilli. Miller et al (36) demonstrated that the post-phagocytic burst of oxygen consumption in polymorphonuclear cells did not occur following ingestion of virulent *S. typhi*. Thus, though ingested, the organisms may not be killed and, indeed, may undergo intracellular multiplication.

Nonspecific Immunity

Usually described as unrelated to humoral or cellular immunity, nonspecific immunity encompasses many different processes, some understood and others poorly defined. Two of the best-studied factors involved in nonspecific defense against intestinal infection are gastric acidity and normal bowel flora. Investigations of nontyphoidal salmonella, shigella, and cholera indicate that gastric resection or gastric hypochlorhydria predispose the individual to greater risk of infection and to more severe illnesses (37). Furthermore, antacids have a marked effect upon the ability of the intestinal tract to limit infection. Simultaneous oral administration of sodium bicarbonate and challenge organisms to volunteers resulted in lower numbers of *V. cholerae* causing infection (38) and of greater fecal excretion of *Shigella flexneri* (39).

The influence of normal bowel flora on the ability of salmonella to successfully colonize the gut was examined in rats. Pretreatment with streptomycin, which killed bacteroides and lactobacilli, allowed the ID₅₀ (the dose of organism causing infection in 50% of the cases) of *Salmonella enteritidis* to fall from a million to less than ten organisms. This was associated with a marked decrease in the presence of short-chain fatty acids, produced by bacteroides and lactobacilli, and a simultaneous rise in pH in the large intestine (40). Similar results were obtained in volunteers. After treatment with streptomycin, four volunteers became ill with typhoid fever after

ingestion of 1000 viable typhoid bacilli, a small dose usually not associated with disease. None of the 14 controls developed disease with this dose (13).

In summary, intestinal factors, both organism-specific such as secretory IgA, and nonspecific such as gastric acidity, intestinal motility, and presence of normal bowel flora, act as the initial mechanisms preventing invasion of the intestinal mucosa. Cell-mediated immunity and phagocytosis undoubtedly play an important role in the lamina propria, mesenteric lymph nodes, and reticuloendothelial system once the mucosa has been penetrated. Serum antibodies against O, H, and Vi antigens are frequently increased following typhoid fever but their importance in immunity is difficult to ascertain.

HISTORY OF TYPHOID VACCINES

The sophisticated knowledge of immunology was long preceded by empirical attempts to elicit immunity against typhoid fever. The typhoid bacillus was first demonstrated in mesenteric lymph nodes by Eberth in 1880. Not long afterward, Pfeiffer (41) isolated *S. typhi* from stool and, in 1896, reported that guinea pigs could be protected from lethal doses of typhoid bacilli by serum from convalescing typhoid fever patients. In that same year, Widal (4) added serial dilutions of serum from patients with typhoid fever to broth cultures of *S. typhi* and observed gross flocculation.

The stage was set for attempts at control of typhoid fever, and, indeed, the first efforts at human vaccination were made in that same year. Almroth Wright, of the British Army Medical School, injected typhoid bacilli, killed with heat and phenol, into two Indian medical officers and a horse. Subsequently, one man ingested viable *S. typhi*; because he did not develop clinical typhoid fever, it was assumed that the vaccination was successful (42). Following this small volunteer trial, Wright conducted a sort of field trial among army units sent to India; the results were inconclusive. Another poorly controlled trial was carried out in 1899 with voluntary vaccinees in army units in the Boer War (43). Again, the results were ambiguous and the issue of typhoid vaccination became a controversial one in Britain. At the onset of World War I in 1914, typhoid vaccination was voluntary in the British Army. However, Wright wrote an emphatic appeal for the use of typhoid vaccine in British troops and, by the end of 1915, almost all of the British expeditionary force had been vaccinated (43).

In the United States, interest in typhoid vaccine was later than that in England and was initially directed at development of an oral vaccine. James Carroll, Walter Reed's colleague in studies of yellow fever, attempted to develop an oral typhoid fever vaccine, but his 1904 experiments were

dramatically unsuccessful (44). He prepared his vaccine by heating a broth culture of *Salmonella typhi* in an Erlenmeyer flask to 56°C for one hour. The flask was then incubated for one week at 36°C after which 1-cc aliquots were cultured to judge sterility. After demonstrating increased antibody levels in six guinea pigs and six rabbits, Carroll appealed to the Surgeon General for volunteers to test the vaccine in humans. On July 12, 1904, the volunteers assembled for the first dose of the vaccine. To assuage any apprehension on the part of the volunteers, Carroll and two medical colleagues joined the ten volunteers in drinking the vaccine preparation. Two days later, the volunteers began a march from Washington, DC, to Gettysburg, Pennsylvania. There, eight of the ten volunteers developed illness consistent with typhoid fever; two of the three investigators developed similar illnesses. Laboratory confirmation was not attempted in many of the men; however, a blood culture of one man contained *S. typhi*, and the sera of two other men were positive for the agglutination test recently introduced by Widal. After this outbreak, the vaccine was reexamined and found to contain 2-3 organisms/cc of broth. Carroll repeated the preparation and again found small numbers of organisms. It is quite likely that the initial heating was inadequate to kill all bacilli in the center of the flask; inadequate refrigeration may have allowed subsequent growth of organisms. It is ironic that this attempt at protection against the typhoid bacillus was the first demonstration of Koch's postulates in regard to typhoid fever.

However, because of the experience with the British army, which demonstrated relatively mild reactions and possible efficacy, parenteral vaccine was introduced to the American army and to the armies of a number of countries involved in World War I. The incidence of typhoid fever in American troops in this war was considerably lower than that of troops in the Spanish American War 15 years earlier. While this was attributed to the efficacy of the vaccine, general sanitation and water purification measures also improved between the two wars. Typhoid fever likewise presented less of a problem in World War II; yet again, the role of the vaccine could not be adequately gauged (43).

Prior to the 1960s, there were few well-controlled studies, i.e. ones demonstrating reasonable certainty that vaccinated and unvaccinated individuals had experienced similar postvaccine and, indeed, prevaccine exposure to the pathogen. One of the few studies that could demonstrate comparable exposure was that of a 1946 outbreak among civilians and military personnel who were exposed to apparently contaminated orange juice (45). Of 211 individuals who had received at least one vaccine dose at any time, only one became infected, whereas, of the 140 unimmunized persons, 17 (12.1%) developed typhoid fever, a significantly higher incidence of the disease.

RECENT TRIALS OF PARENTERAL VACCINES

Vaccine efficacy was not evaluated prospectively until well-controlled trials were conducted in the 1960s. These trials, under the auspices of the World Health Organization (WHO), were conducted in Yugoslavia (7), Guyana (8, 9), and Poland (10); similar trials were carried out in the Soviet Union (46). These were areas of the world where typhoid was endemic and typhoid fever was a disease of childhood, many adults having gained immunity presumably from one or more unrecognized, perhaps subclinical, infections.

Two of the most effective vaccines used in these trials were developed at Walter Reed Army Institute of Research from a seed batch of *S. typhi*. One half of the batch was killed by the time-honored method of heating to 56°C for one hour, then treated and suspended in phenol; this was designated vaccine L. The other half of the batch was inactivated by suspension in acetone and was coded K. All trials were prospective and the vaccine groups, as well as the control group (the latter received tetanus toxoid), were assigned by methods intended to assure comparability among the groups. The demographic characteristics of each group within a trial were equivalent, which indicated adequate controls. Diagnosis was made by isolation of *S. typhi*, generally from blood or, with an appropriate history, from feces or urine; the Widal test was not used in the diagnosis.

The results of these field trials, which involved large numbers of persons, indicated that both of these vaccines provided significant immunity, and that vaccine K, the acetone-inactivated vaccine, was the more successful of the two. Vaccine K proved 79–94% effective over the various periods of follow-up (47) (See Table 1). The trial in Guyana examined only school children under 15 years of age and had the longest follow-up: It demonstrated that two doses of the vaccine, given four weeks apart, resulted in 88% protection even at seven years of follow-up (9).

A finding in the Yugoslavian studies may have demonstrated one of the difficulties in running field trials with part of a population voluntarily participating. The control group, randomly assigned to receive the tetanus toxoid vaccine, had a typhoid fever rate of 12.9 per 1000. An examination of the population that did not participate in the trial at all and thus received neither typhoid nor tetanus toxoid vaccine, revealed an attack rate of 81.8 per 1000, more than six times the rate of the control group. A possible reason for this difference between the control participants and the nonparticipants may be that the tetanus toxoid conferred some nonspecific immunity to the study controls not elicited in the unvaccinated part of the population. More likely, however, is the possibility that those people interested enough to volunteer for a randomized study might have been from a

Table 1 Results of selected typhoid vaccine trials

	Country	Reference	Type study	Age com- position	Followup (years)	Vaccine group		Control group		Vaccine efficacy (%) ^a
						Cases of typhoid	Persons in trial	Cases of typhoid	Persons in trial	
PARENTERAL										
Acetone-inactivated	Yugoslavia	7	Field trial	Adults and children	3	16	5,028	75	5,039	79
	Guyana	8	Field trial	Children	4	6	24,046	99	24,241	94
	Poland	10	Field trial	Children	2	4	81,534	31	83,734	87
	Tonga	48	Field trial	Adults and children	3 ^b	13	11,128	26	11,129	50
	US	13	Volunteer challenge	Adults	≤1	4	43	28	104	67
ORAL										
Epimerase-less	US	15	Volunteer challenge	Adults	<1	2	28	23	43	87

^a Efficacy = $100(b-a)/b$ where a indicates incidence in vaccinated group and b is incidence in control group.^b In order to compare with other field trials, the incidence during the first three years of study is shown.

part of the population with sufficient health concern and knowledge to have practiced general hygiene on a level above that of the nonparticipants.

Another interesting and unexpected finding occurred in Guyana because a small proportion of persons did not return for a second dose. Those individuals had immunity at least equal to the participants who received both doses. Similar findings were reported from Poland (47). However, because the groups receiving only one dose were small and were not randomly selected, another trial to test the efficacy of the single-dose regimen was conducted in Tonga in 1966-1973 (48). In this country, one dose of acetone-inactivated vaccine conferred only about one year's protection as compared to the four-to-five years' protection offered by the two-dose regimen. Moreover, the Tongan field trial was interesting in that, albeit significantly greater than the control, the rate of protection offered by the acetone-inactivated vaccine (about 50% at three years) was lower than that found in the earlier field trials in other parts of the world (See Table 1). The investigators suggested that in Tonga a prominent route of transmission is food where higher concentrations of organisms might be present. In the countries involved in the earlier studies, water-borne transmission, with possibly smaller concentrations of organisms, was the major route of transmission.

Volunteer studies at the University of Maryland offered good evidence for such an explanation. In these studies, also testing vaccines K and L, the exact challenge doses of viable *S. typhi* were measured. The vaccinees, given three doses of the respective vaccines, rather than two as in the WHO studies, were challenged from three months to one year after vaccination. When an oral dose of 10^9 viable typhoid bacilli was administered, the vaccine provided no protection. Similarly, ingestion of 10^7 organisms overcame any effect of the vaccine. However, when 10^5 bacilli were swallowed, both K and L provided 67% protection (13) (See Table 1). One can assume that the water-borne method of transmission provides lower inocula of *S. typhi* than does the food-borne method in which organisms may have multiplied for several hours after contamination. The volunteer studies suggest that the parenteral vaccine would exhibit relatively good protection if exposure were to relatively low numbers of organisms but the vaccine effect would be overwhelmed with higher inocula.

Furthermore, the volunteer studies demonstrated vaccine efficacy in a population not previously exposed to natural infection by virtue of residence in endemic areas. The Yugoslavian field trials revealed a higher attack rate in children than in adults (7). This finding suggests that age or the number of years in an endemic area is associated with resistance to the disease, most likely induced immunity from subclinical infections. Consequently, vaccine trials in American volunteers is of considerable importance in demonstrat-

ing vaccine efficacy without benefit of natural immunizing experience with *S. typhi*. Similarly, the relatively low protection rate of 67% in the American volunteers may be attributed to the absence of previous exposure to *S. typhi*, so that an anamnestic response was not elicited.

ORAL VACCINE

As mentioned earlier in the description of Carroll's unfortunate experience with an inadequately killed vaccine preparation, attempts at oral vaccination against typhoid fever have a history almost as long as that of parenteral vaccines. Indeed, oral vaccines have been commercially available for 50 years. Their effectiveness, however, has been questioned even more than that of the parenteral vaccines. The pursuit of such vaccines has been intensified coincident with the success of oral poliomyelitis vaccine and the recent interest and gains in knowledge of gastrointestinal immunity. The advantages that might be offered by an oral vaccine over the present parenteral vaccine are practical: greater ease of administration, fewer systemic reactions, and, one would hope, better protection.

Certainly the fact that the typhoid bacillus has a defined site of entry at the intestinal mucosa makes that surface an obvious target for the establishment of effective barriers. The natural history of the disease implies the existence of defense mechanisms that restrict organisms to the intestinal lumen, and prevent attachment and invasion of the mucosal surface. Certainly chronic carriers of typhoid fever can shed large numbers of organisms (as many as 10^{10} /gm feces) without evidence of symptomatic disease (49). The continued exposure without evident illness of persons living in endemic locales with poor sanitation suggests that repeated exposures to subinfectious inocula may elicit intestinal immunity, which prevents subsequent disease. Oral vaccines could be administered to duplicate this type of naturally acquired resistance.

One of the controversial issues concerning oral vaccines is the relative efficacy of live and killed vaccines (50). The proponents of live, attenuated vaccines point to the success of the oral poliomyelitis vaccine. Their major points are (a) live vaccine is more likely to contain the essential antigen(s) necessary to elicit adequate immunity than is a killed vaccine, because the processes potent enough to kill the organism may be more likely to destroy this antigen(s), and (b) multiplication of live organisms can take place, which thus allows a greater, and perhaps more prolonged, antigenic stimulus. The killed vaccine proponents, on the other hand, warn of the danger of live organisms reverting to the wild state. On the positive side, they point to the possibility of separating the antigen and administering it as a molecular vaccine with potentially more successful immunization and fewer reac-

tions. As described above, however, the identification, much less isolation, of antigens of *Salmonella typhi* that are important to immunogenicity is a feat that has eluded investigators for decades. Consequently, as with the parenteral vaccines, the development of oral vaccines has been an empiric process. Also like the parenteral typhoid vaccine, controlled trials examining efficacy have only recently been performed. At this point, only killed oral vaccines have been examined in field trials. The best of these controlled trials were conducted in India and demonstrated no beneficial effect (51, 52). Less-well-designed trials were administered in Chile (53) and the USSR (54); they also showed poor vaccine efficacy. The volunteer studies in the US showed similarly poor results with killed oral vaccine (55).

However, recent investigations of live oral vaccine against typhoid fever in humans have been conducted and are quite promising. The most encouraging studies employed a strain isolated by Germanier & Furer of the Swiss Serum Institute (56). This organism, labeled Ty21a, lacks the enzyme uridine-diphosphate-glucose-4-epimerase. In the absence of an external supply of galactose, these strains cannot form UDP-galactose and cannot incorporate galactose into the lipopolysaccharide of the cell wall. Consequently, the strain has incompletely developed side chains and lacks O antigens. In the presence of galactose, this strain can form UDP-galactose, but accumulation of this molecule and of galactose-1-phosphate results in osmotic lysis of the cell. Therefore, growth in galactose-containing medium may result in greater immunogenicity because of enhanced production of O antigen, while low virulence is maintained because of the osmotic lysis induced.

American volunteers were given the galactose-grown strain in milk after two grams of NaHCO_3 were ingested; this administration was repeated 5–8 times over a 4-week period (15). Within 5–9 weeks after completion of the course of vaccination, the vaccinees as well as a group of unvaccinated volunteer controls were challenged with 10^5 virulent *S. typhi*. Of the participants who received the galactose-grown vaccine, only 7% became ill with typhoid fever compared to 53% of the unvaccinated controls, a difference that was highly significant ($P = .0002$) and that demonstrated a vaccine efficacy of 87% (See Table 1). Not only was a high degree of protection from clinical disease achieved in these people, but a much smaller proportion of them (11%) excreted virulent typhoid bacilli after three days as compared with the controls (60%) ($P = .00009$). Side effects from the vaccine were mild; no placebo was given to the controls for comparison. Among 958 stool isolates taken during the study, none reverted to the wild type.

The epimerase-less strain achieved 87% effectiveness as a live oral vaccine, the highest of any typhoid vaccine, parenteral or oral, yet studied at the University of Maryland. Of additional importance was the shorter duration of fecal excretion of the virulent *S. typhi* following challenge. This

undoubtedly was related to intestinal defenses and may be important in limiting spread of typhoid fever. It should be stressed, however, that up to eight oral doses of vaccine were administered and that the challenge was given soon after vaccination. A field trial currently underway in Egypt may help to determine the duration and magnitude of protection and the effect of a smaller number of doses and of lyophilization upon vaccine effectiveness. One might expect that the excellent results obtained in American volunteers would be improved in individuals from endemic areas by boosting partial immunity already present from previous natural exposure.

As a caveat, we must mention one aspect of the disease that casts a shadow over the field of typhoid vaccine development: Immunity following actual infection with *S. typhi* may be only partial. Marmion et al (57) reported two typhoid fever epidemics, caused by different strains of *S. typhi*, in a British Air Force base in Egypt within a five-month period in 1950. Fifty-four men who contracted typhoid fever during the first outbreak were exposed during the second. The incidence of disease among these men during the second epidemic was 20% compared with 31% among newcomers to the base who were unexposed during the first outbreak. DuPont et al (55), studying reexposure of volunteers to the same organism and in the same dose, obtained almost identical attack rates as found in the outbreaks reported by Marmion; 23% among the previously infected vs 30% among the volunteers never before exposed. A number of interpretations of these data can be offered: (a) those persons becoming ill for the second time may have had one or another unidentified defect in immune defense; (b) treatment of the first infection with chloramphenicol may have blunted the immune response; or (c) the important point to consider is not that 20% became infected but that 80% appeared to be protected. The actual explanation for these phenomena is not available at present, so we must accept the implication that achieving a high level of immunity following vaccination may be a difficult task.

SUMMARY AND RECOMMENDATIONS

In the absence of excellent public hygiene and sanitation on a worldwide scale, the need for an effective vaccine against typhoid fever persists. The continuing success of oral polio vaccine and the increasing interest and knowledge of gastrointestinal immunity have prompted recent investigations into oral vaccines against typhoid fever. To date, no properly studied oral vaccine containing killed organisms has provoked adequate immunity. However, live but avirulent mutants of *S. typhi* show promise. Perhaps the most encouraging live oral vaccine has been an epimerase-less mutant of *S. typhi*. Volunteers ingesting multiple doses of this avirulent organism

demonstrated 87% immunity at the time of challenge five to nine weeks later. Further clinical investigation is needed to determine the value of this preparation.

An effective oral vaccine is not commercially available at present, so clinical practice is restricted to parenteral vaccines. Recent well-controlled field trials in endemic areas indicate that acetone-inactivated *S. typhi*, parenterally administered as a vaccine, can provide good immunity (79–94%) for three years. Indeed, trials in Guyana demonstrated 88% immunity at the end of seven years. This immunity may be incomplete and may be overwhelmed by large numbers of ingested organisms, however.

The acetone-inactivated vaccine can now be obtained in the United States and should be chosen over the other preparations. Like most previous typhoid vaccines, the acetone-inactivated vaccine causes a relatively high proportion of side effects, primarily local inflammatory reactions and fever; these can be partially controlled with aspirin. The vaccine should be administered subcutaneously and should be repeated once at four weeks. Recurrent or continuous exposure should prompt booster doses every three years. The addition of paratyphoid A and B vaccine is of unproven benefit in prevention of these diseases and should be avoided. Persons who should receive the vaccine are those who are traveling to or living in areas endemic for the disease, particularly in the villages or smaller cities, those persons working with the live organism in laboratories, and those persons involved in outbreaks of typhoid fever. Because of the extremely low prevalence of the disease and of carriers in the United States, it should rarely be necessary in this country to suggest mass vaccination against typhoid fever even in such disasters as floods.

Literature Cited

1. Means, J. H. 1960. Reflections on typhoid fever. *King's Coll. Hosp. Gaz.* 39:270–78
2. Baine, W. B., Farmer, J. J. III, Gangarosa, E. J., Hermann, G. T., Thornsberry, C., Rice, P. A. 1977. Typhoid fever in the United States associated with the 1972–1973 epidemic in Mexico. *J. Infect. Dis.* 135:649–53
3. Olarte, J., Galindo, E. 1973. *Salmonella typhi* resistant to chloramphenicol, ampicillin, and other antimicrobial agents: strains isolated during an extensive typhoid epidemic in Mexico. *Antimicrob. Agents Chemother.* 4:597–601
4. Widal, F. 1896. Sero diagnostic de la fièvre typhoïde. *Sem. Med.* 16:259
5. Benenson, A. S. 1964. Serological responses of man to typhoid vaccines. *Bull. WHO* 30:653–62
6. Yugoslav Typhoid Commission. 1962. A controlled field trial of the effectiveness of phenol and alcohol typhoid vaccines. *Bull. WHO* 26:357–69
7. Yugoslav Typhoid Commission. 1964. A controlled field trial of the effectiveness of acetone-dried and inactivated and heat-phenol-inactivated typhoid vaccines in Yugoslavia. *Bull. WHO* 30:623–30
8. Ashcroft, M. T., Morrison Ritchie, J., Nicholson, C. C. 1964. Controlled field trial in British Guiana school children of heat-killed-phenolized and acetone-killed -lyophilized typhoid vaccines. *Am. J. Hyg.* 79:196–206
9. Ashcroft, M. T., Singh, B., Nicholson, C. C., Ritchie, J. M., Sobryan, E., Williams, F. 1967. A seven-year field trial

- of two typhoid vaccines in Guyana. *Lancet* 2:1056-59
10. Polish Typhoid Committee. 1966. Controlled field trials and laboratory studies on the effectiveness of typhoid vaccines in Poland, 1961-64. *Bull. WHO* 34: 211-22
 11. Hornick, R. B., Woodward, T. E. 1966. Appraisal of typhoid vaccine in experimentally infected human subjects. *Trans. Am. Clin. Climatol. Assoc.* 78: 70-78
 12. Hornick, R. B., Woodward, T. E., McCrumb, F. R., Snyder, M. J., Dawkins, A. T., Bulkeley, J. T., de la Macorra, F., Carozza, F. A. 1967. Typhoid fever vaccine—yes or no? *Med. Clin. North Am.* 51:617-23
 13. Hornick, R. B., Greisman, S. E., Woodward, T. E., DuPont, H. L., Dawkins, A. T., Snyder, M. J. 1970. Typhoid fever: pathogenesis and immunologic control. *N. Engl. J. Med.* 283:686-91
 14. Joo, I. 1970. Present status and perspectives of vaccination against typhoid fever. Presented at the Int. Conf. Appl. Vaccines against Viral, Rickettsial, Bacterial Dis. Man. *Pan Am. Health Organ. Washington, DC*, pp. 329-41
 15. Gilman, R. H., Hornick, R. B., Woodward, W. E., DuPont, H. L., Snyder, M. J., Levine, M. M., Libonati, J. P. 1977. Evaluation of a UDP-glucose-4-epimeraseless mutant of *Salmonella typhi* as a live oral vaccine. *J. Infect. Dis.* 136: 717-23
 16. Levine, M. M., DuPont, H. L., Hornick, R. B., Snyder, M. J., Woodward, W., Gilman, R. H., Libonati, J. P. 1976. Attenuated, streptomycin-dependent *Salmonella typhi* oral vaccine: potential deleterious effects of lyophilization. *J. Infect. Dis.* 133:424-29
 17. Wahdan, M. H., Sippel, J. E., Mikhail, I. A., Rahka, A. E., Anderson, E. S., Sparks, H. A., Cvjetanovic, B. 1975. Controlled field trial of a typhoid vaccine prepared with a nonmotile mutant of *Salmonella typhi* Ty2. *Bull. WHO* 52:69-73
 18. Tully, J. G., Gaines, S., Tigertt, W. D. 1963. Studies on infection and immunity in experimental typhoid fever. IV. Role of H antigen in protection. *J. Infect. Dis.* 112:118-24
 19. Guentzel, M. N., Berry, L. J. 1975. Motility as a virulence factor for *Vibrio cholerae*. *Infect. Immun.* 11:890-97
 20. Walker, W. A., Isselbacher, K. J. 1977. Intestinal antibodies. *N. Engl. J. Med.* 297:767-73
 21. Heremans, J. F., Vaerman, J.-P. 1971. Biological significance of IgA antibodies in serum and secretions. *Prog. Immun. Proc. 1st Int. Congr. Immunol.*, pp. 875-90
 22. Oral enteric bacterial vaccines. 1972. *WHO Tech. Rep. Ser. No. 500* 34 pp.
 23. Girard, J. P., de Kalbermatten, A. 1970. Antibody activity in human duodenal fluids. *Eur. J. Clin. Invest.* 1:188-95
 24. Crabke, P. A., Carbonara, A. O., Heremans, J. F. 1965. The normal human intestinal mucosa as a major source of plasma cells containing Gamma A-immunoglobulin. *Lab. Invest.* 14: 235-43
 25. Svennerholm, A.-M., Lange, S., Holmgren, J. 1978. Correlation between intestinal synthesis of specific immunoglobulin A and protection against experimental cholera in mice. *Infect. Immunol.* 21:1-6
 26. Williams, R. C., Gibbons, R. J. 1972. Inhibition of bacterial adherence by secretory immunoglobulin A: a mechanism of antigen disposal. *Science* 177: 697-99
 27. Bienenstock, J. 1974. The physiology of the local immune response and the gastrointestinal tract. *Prog. Immunol.* 4: 197-207
 28. Fuburg, E. S., Freter, R. 1973. Protection against enteric bacterial infections by SIgA. *J. Immunol.* 111:395-403
 29. McClelland, D. B. L., Sampson, R. R., Parkin, D. M. 1972. Bacterial agglutination studies with secretory IgA prepared from human gastrointestinal secretions and colostrum. *Gut* 13: 450-58
 30. Muller-Schoop, J. W., Good, R. A. 1975. Functional studies of Peyer's patches: evidence for their participation in intestinal immune responses. *J. Immunol.* 114:1757
 31. Sprinz, H., Gangarosa, E. J., Williams, M., Hornick, R. B., Woodward, T. E. 1966. Histopathology of the upper small intestines in typhoid fever. *Am. J. Dig. Dis.* 11:615-24
 32. Balakrishna Sarma, V. N., Malaviya, A. N., Kumar, R., Ghai, O. P., Bakhtary, M. M. 1977. Development of immune response during typhoid fever in man. *J. Exp. Immunol.* 28:35-39
 33. Blander, R. V., Mackaness, G. B., Collins, F. M. 1966. Mechanisms of acquired resistance in mouse typhoid. *J. Exp. Med.* 124:573-600
 34. Wagner, H. N., Iio, M., Hornick, R. B. 1963. Studies of the reticuloendothelial

- system. II. Changes in the phagocytic capacity of the RES in patients with certain infections. *J. Clin. Invest.* 42: 427-34
35. Morello, J. A., Baker, E. E. 1965. Interaction of salmonella with phagocytes *in vitro*. *J. Infect. Dis.* 115:131-41
 36. Miller, R. M., Garbus, J., Hornick, R. B. 1972. Lack of enhanced oxygen consumption by polymorphonuclear leukocytes on phagocytosis of virulent *Salmonella typhi*. *Science* 175:1010-11
 37. Giannella, R. A., Broitman, S. A., Zameck, N. 1973. Influence of gastric acidity on bacterial and parasitic enteric infections. *Ann. Intern. Med.* 78:271-76
 38. Hornick, R. B., Music, S. I., Wenzel, R., Cash, R., Libonati, J. P., Snyder, M. J., Woodward, T. E. 1971. The Broad Street pump revisited: response of volunteers to ingested cholera vibrios. *Bull. NY Acad. Med.* 47:1181-91
 39. DuPont, H. L., Hornick, R. B., Snyder, M. J., Libonati, J. P., Formal, S. B., Gangarosa, E. J. 1972. Immunity in shigellosis. I. Response of man to attenuated strains of shigella. *J. Infect. Dis.* 125:5-11
 40. Hentges, D. J., Maier, B. R. 1970. Inhibition of *Shigella flexneri* by the normal intestinal flora. III. Interactions with *Bacteroides fragilis* strains *in vitro*. *Infect. Immun.* 2:364-70
 41. Pfeiffer, R., Kolle, W. 1896. Ueber die spezifische immunitätsreaction der Typhus-bacillen. *Z. Hyg. Infektionskr.* 21:203-46
 42. Wright, A. E., Semple, D. 1897. Remarks on vaccination against typhoid fever. *Br. Med. J.* 1:256-59
 43. Ashcroft, M. T. 1963. Immunization against typhoid fever. *West Indian Med. J.* XII:82-89
 44. Tigert, W. D. 1959. The initial effort to immunize American soldier volunteers with typhoid vaccine. *Mil. Med.* 124: 342-49
 45. Duncan, T. G., Doull, J. A., Miller, E. R., Bancroft, H. 1946. Outbreak of typhoid fever with orange juice as the vehicle, illustrating the value of immunization. *Am. J. Public Health* 36:34-36
 46. Hejfec, L. B., Salmin, L. V., Lejtman, M. Z., Kuz'minova, M. L., Vasil'eva, A. V., Levina, L. A., Bencianova, T. G., Pavlova, E. A., Antonova, A. A. 1966. A controlled field trial and laboratory study of five typhoid vaccines in the USSR. *Bull. WHO* 34:321-39
 47. Cvjetanovic, B., Uemura, K. 1965. The present status of field and laboratory studies of typhoid and paratyphoid vaccines. *Bull. WHO* 32:29-36
 48. Tapa, S., Cvjetanovic, B. 1975. Controlled field trial on the effectiveness of one and two doses of acetone-inactivated and dried typhoid vaccine. *Bull. WHO* 52:75-80
 49. Merselis, J. G. Jr., Kaye, D., Connolly, C. S., Hook, E. W. 1964. Quantitative bacteriology of the typhoid carrier state. *Am. J. Trop. Med. Hyg.* 13:425-29
 50. Edsall, G. 1975. Live versus killed bacterial vaccines. *Monogr. Allergy* 9: 231-44
 51. Chuttani, C. S., Prakash, K., Vergese, A., Gupta, P., Chawla, R. K., Grover, V., Agarwal, D. S. 1973. Ineffectiveness of an oral killed typhoid vaccine in a field trial. *Bull. WHO* 48:756-57
 52. Chuttani, C. S. 1976. Controlled field trials of three different oral killed typhoid vaccines in India. *Dev. Biol. Stand.* 33:98-101
 53. Borgono, J. M., Corey, G., Engelhardt, H. 1976. Field trials with killed oral typhoid vaccines. *Dev. Biol. Stand.* 33: 80-84
 54. Hejfec, L. B., Levina, L. A., Salmin, L. B., Antonova, A. A., Segal, L. S., Kuz'minova, M. L., Slavina, Kh. M., Vasileva, A. V. 1976. Controlled field trials of killed oral typhoid and paratyphoid B vaccines and cell-free, chemical aerosol typhoid vaccine. *Dev. Biol. Stand.* 33:93-97
 55. DuPont, H. L., Hornick, R. B., Snyder, M. J., Dawkins, A. T., Heiner, G. G., Woodward, T. E. 1971. Studies of immunity in typhoid fever. Protection induced by killed oral antigens or by primary infection. *Bull. WHO* 44:667-72
 56. Germanier, R., Furer, E. 1975. Isolation and characterization of *Gal* E mutant Ty 21a of *Salmonella typhi*: A candidate strain for a live, oral typhoid vaccine. *J. Infect. Dis.* 131:553-58
 57. Marmion, D. E., Naylor, G. R. E., Stewart, I. O. 1953. Second attacks of typhoid fever. *J. Hyg.* 51:260-67

ENDOCRINOLOGY OF THE PLACENTA

◆7336

Royce B. Everett, M.D. and Paul C. MacDonald, M.D.

Cecil H. and Ida Green Center for Reproductive Biology Sciences
and the Departments of Obstetrics and Gynecology and Biochemistry,
University of Texas, Southwestern Medical School, Dallas, Texas 75235

INTRODUCTION

The endocrine physiology of pregnancy is of great interest and considerable wonderment. The quantity of steroid and protein hormones produced during normal human pregnancy staggers the imagination; and the physiologic adaptations of women during pregnancy are among the most remarkable modifications that adult humans undergo. During the last few weeks of pregnancy, 15 mg of estradiol (E2) (1), 60 mg of estriol (E3) (2), and 250 mg of progesterone (3) are secreted daily by the placenta into the maternal vasculature.

Early in the modern era of steroid hormone investigations, it was recognized that massive amounts of progesterone and estrogen were produced during human pregnancy. But only in the last two decades have we learned of the unique mechanisms by which steroid hormones are formed in the human placenta. Prior to that time, it was believed that the placenta synthesized steroid hormones *de novo*, i.e. from acetate, as do other endocrine organs such as the gonads and the adrenal cortex. However, a number of investigators demonstrated unequivocally that the placenta, with respect to steroid biosynthesis, is an incomplete endocrine organ, that is the placenta depends upon circulating precursors for its steroid hormone production.

Large quantities of protein hormones, similar but not identical to those produced by the anterior pituitary, have also been isolated from the blood and urine of pregnant women and from chorionic tissue. These include human chorionic gonadotropin (hCG), human placental lactogen (hPL), and a protein hormone with thyrotropic properties similar to those of

pituitary thyroid-stimulating hormone, i.e. chorionic thyrotropin. Some investigators have suggested that a chorionic ACTH is also produced by the placenta. To bring the placenta to an even more interesting endocrine posture, it has been demonstrated that trophoblasts also produce luteinizing-hormone-releasing hormone (LHRH) (4).

PROGESTERONE

Progesterone likely plays many roles in human gestation, some of which remain enigmatic. However, progesterone is *the* hormone of pregnancy and is of undoubted importance in the maintenance of early pregnancy, of probable importance in the control of Na^+ excretion by the kidneys of pregnant women, and of presumed importance in the events of parturition. Results of recent studies provide support for the view that progesterone serves an important role in the regulation of maternal blood pressure during human pregnancy (5). In many pregnant mammals, progesterone is formed primarily in the corpus luteum, but during human gestation progesterone production by the corpus luteum is supplanted by the placental formation of progesterone after approximately six to nine weeks of gestation (6). The rate of progesterone secretion during human gestation appears to be proportional to placental mass; indeed, the rate of progesterone secretion is not markedly altered by changes in fetal health or by fetal abnormalities. This fact was dramatically demonstrated by Cassmer (7), who found that following ligation of the umbilical cord in pregnancies during the second trimester, there was no change in the rate of maternal urinary pregnanediol excretion for several days. While the results of these studies provided clear evidence that the fetus was not involved in placental progesterone biosynthesis, it was not known at that time that the placenta was incapable of *de novo* synthesis of progesterone, i.e. the placenta does not possess the enzymes necessary to carry out the entire series of steps in steroid hormone biosynthesis beginning with acetate. In an elegant study, Hellig et al (8) demonstrated that the placenta utilizes cholesterol assimilated from the maternal circulation for progesterone synthesis. These investigators found that when radiolabeled cholesterol was administered to women pregnant with anencephalic fetuses, within two days the specific activity of progesterone and urinary pregnanediol became the same as that of plasma cholesterol, demonstrating that maternal plasma cholesterol was the principal precursor of placental progesterone synthesis.

Recently the cellular mechanisms of the trophoblastic uptake of maternal plasma cholesterol have been clarified. Simpson and co-workers (9, 10) found that choriocarcinoma cells in culture preferentially utilize cholesterol derived from low density lipoprotein (LDL) for pregnenolone and progesterone synthesis.

terone biosynthesis. Goldstein & Brown (cf 11), working with human fibroblasts in culture, provided new insights into the mechanisms of cellular assimilation of cholesterol from plasma lipoproteins. Their investigations showed that many nonhepatic human tissues take up cholesterol through a process that involves binding of LDL to specific high affinity plasma membrane receptors. LDL is internalized by an adsorptive endocytotic process, and the internalized vesicles fuse with lysosomes. Lysosomal enzymes effect the hydrolysis of the protein component of LDL, giving rise to free amino acids; hydrolysis of the cholesterol ester component of LDL gives rise to free cholesterol and free fatty acids. In the case of the trophoblasts, free cholesterol, derived from LDL, is available for progesterone biosynthesis (10). The free cholesterol also inhibits the action of the enzyme HMG-CoA-reductase, the rate-limiting step in the cellular synthesis of cholesterol from 2-carbon fragments (10). Thus, it can be reasoned that the rate of placental progesterone synthesis is dependent upon the content of LDL receptors on the plasma membranes of the trophoblasts; therefore, the rate of progesterone biosynthesis is principally independent of the rate of maternal placental blood flow.

Of the daily placental production of progesterone, most (perhaps 90% or more) enters the maternal compartment (12). Even so, the levels of progesterone in umbilical vein plasma are very high, attaining levels of 600 ng/ml or more (13).

It has been demonstrated that fetal plasma progesterone can be utilized by the fetal adrenals to produce corticosteroids (14). Interestingly, the activity of 3β -hydroxysteroid dehydrogenase in the fetal adrenal is reduced—likely by suppression through the action of estrogen and progesterone that reach the fetus from the placenta. Thus, maternal plasma cholesterol, after conversion to progesterone in the placenta, may constitute a precursor for fetal adrenal corticosteroid synthesis. The role of fetal adrenal cortisol in the process of fetal lung maturation is of urgent concern and constitutes an interesting area of fetal physiology currently under intense study in many laboratories.

The importance of progesterone withdrawal in the initiation of parturition has been demonstrated in several animal models. In such species a fall in maternal plasma progesterone levels heralds the onset of labor (15). This phenomenon, however, is not consistently observed prior to human parturition (16). However, one model for human parturition envisions that lysosomes in the fetal membranes are stabilized by the action of progesterone, and that alterations in the intracellular concentration of progesterone may bring about labilization of lysosomes; this leads to expression of lysosomal enzyme activity, which leads to a sequence of events that culminate in prostaglandin formation, myometrial contractions, and labor (17).

Unfortunately, knowledge of the levels of maternal plasma progesterone concentrations in human pregnancy is of little or no utility in evaluating fetal well-being or in selecting optimal clinical management, including the timing of delivery of pregnancies in which the fetus is believed to be at risk.

ESTROGEN

Early in the study of steroid hormones it was recognized that pregnancy was associated with strikingly elevated levels of estrogens. However, the production of the three classic estrogens—E1, E2, and E3—in pregnant women was found to differ not only quantitatively but qualitatively from that of the nonpregnant state. Specifically, during human pregnancy there is a disproportionate rise in the production of E3 as compared to that of E1 and E2. In nonpregnant women, the ratio of urinary E3 to the sum of urinary E1 and E2 is approximately one. This ratio, however, increases during pregnancy such that the level of urinary E3 becomes approximately ten times greater than the sum of urinary E1 and E2. It is now known that this disproportionate rise in E3 is accounted for by an independent pathway for E3 biosynthesis in the placenta during human pregnancy (1).

The commonly held view that the placenta synthesized estrogens *de novo* was seriously challenged first by Frandsen & Stakeman (18), who found that the rate of production of urinary estrogens in women pregnant with anencephalic fetuses was strikingly less than that in women pregnant with normal fetuses. Since anencephalic fetuses lack the fetal zone of the adrenal cortex, a structure characteristic of normal human fetuses, Frandsen & Stakeman reasoned that the human fetal adrenal supplied the precursors of placental estrogen biosynthesis. In the studies of Cassmer (7), who found that there was no change in maternal urinary pregnanediol excretion, there was a precipitous decrease in estrogen excretion after ligation of the umbilical cord, even though the placenta was left *in situ*. These findings were supportive of the hypothesis proposed by Frandsen & Stakeman, i.e. that the fetal adrenal is a source of placental estrogen precursors. Subsequently, it was demonstrated that the fetal adrenal cortex was not the only source of precursor for placental estrogen synthesis in as much as women pregnant with a hydatidiform mole also produce large quantities of estrogen compared to that produced by nonpregnant persons (19). From the results of these and other studies, it was shown that placental estrogen precursors are also provided by the secretions of maternal adrenal cortex.

The most likely candidate for the role of plasma estrogen precursor for the human placenta is dehydroisoandrosterone sulfate (DS), since this compound is secreted by the normal adrenal cortex, is present in large quantities in the human circulation, and the enzymes required for converting DS to

E2 or to E1 are known to be present in the human placenta. An experimental design in which [^{14}C]DS and [^3H]E2 were administered intravenously to pregnant women showed that a sizeable fraction of maternal plasma [^{14}C]DS was converted to [^{14}C]E2 by the placenta (1, 20). The fractional conversion of maternal plasma DS to E2 increased from about 2–4% during the first trimester of pregnancy to as much as 40–50% in normal pregnant women near term (20). By comparing the fractional conversion of maternal plasma DS to E2 in women at various stages of pregnancy, it appears that the extent of this conversion is primarily dependent upon the rate of delivery of precursor to placenta, i.e. the placental surface area and maternal placental blood flow. In women pregnant with an anencephalic fetus, the production rate of E2 was found to be 25–50% of that found in normal pregnancies; moreover, in women with anencephalic fetuses, estrogen synthesis was derived almost totally from the utilization of maternal plasma DS (21). When maternal adrenal DS secretion was increased by ACTH treatment or decreased by dexamethasone treatment of women pregnant with anencephalic fetuses, there followed a proportional change in the rate of E2 secretion (21). Together, these data support the conclusion that estrogen synthesis by the human placenta is dependent upon DS supplied from the fetal and maternal circulations. These findings were substantiated by the findings of Diczfalussy & Mancuso (22), who also demonstrated, by placental perfusion studies of mid-trimester pregnancies, that fetal and maternal plasma DS accounted for approximately equal amounts of precursor utilized for the secretion of E2 by the placenta.

Major alterations in the metabolic clearance rate (MCR) of maternal plasma DS occur during the course of human pregnancy. Indeed, the MCR-DS increases 5- to 25-fold during human pregnancy, and this increase is primarily the consequence of two rapid routes of clearance of DS from the maternal circulation. The first of these is through the placental conversion of maternal plasma DS to E2, and the second is through 16 α -hydroxylation of DS in the maternal compartment (23). These two routes of removal of DS account for 60–90% of the total MCR of maternal plasma DS in normal gravidas near term. It is important to note that the clearance of DS through the formation of E2 is unique to the placenta, and in late human pregnancy this pathway accounts for 25–50% of the MCR of maternal plasma DS. It is presumed that the clearance of DS to E2, which occurs uniquely in the placenta, is a reflection of the rate of maternal placental perfusion of the placenta, i.e. maternal placental blood flow. This obtains since seemingly only three factors alter the rate of DS conversion to E2 within the placenta. These factors are as follows: (a) the rate of delivery of maternal plasma DS to the placenta, i.e. maternal placental blood flow; (b) the uptake of DS by the trophoblast; and (c) the activities of enzymes catalyzing the conversion

of DS to E2. Since placental uptake of DS appears to depend on simple diffusional transport (21), and since deficiencies of enzymes in the trophoblasts that catalyze the conversion of DS to E2 are rare (24), it has been proposed that alterations in placental perfusion result in a proportional change in the placental clearance of DS to E2 (PC-DSE2). Although measurements of PC-DSE2 are not practical for clinical evaluations of high risk pregnancies, such measurements have been conducted in order to evaluate placental perfusion in complicated pregnancies and to evaluate the physiologic effects of drug treatment of pregnant women. The MCR-DS and the PC-DSE2 have been measured in women with normal and abnormal pregnancies (25-33). In women with pregnancies complicated by preeclampsia and in pregnant women treated with diuretics, a striking reduction in PC-DSE2 was found compared to that observed in normal untreated pregnant women of similar gestational duration. On the other hand, in women with pregnancies in which one would expect greater placental blood flow, e.g. pregnancies with twin fetuses, the PC-DSE2 was higher than normal. The PC-DSE2 is a useful tool for investigating the nature of the physiologic aberrations of pregnancy and for evaluating the effect of various therapeutic modalities on maternal placental perfusion.

Measurements of E3 levels in maternal plasma or urine as a means of evaluating fetal well-being during human pregnancy have fascinated obstetricians for many years. The increased production of E2 by the placenta cannot account for the disproportionate increase in urinary E3 excretion observed during normal human pregnancy. However, an alternate route of synthesis of E3 in pregnancy was suggested when Ryan (34) demonstrated that 16 α -hydroxy-C₁₉-steroids were aromatized by placental tissue yielding E3. The placenta does not possess the enzyme 16 α -hydroxylase. Thus, circulating 16 α -hydroxy-C₁₉-steroids were considered to be likely precursors of placental E3 synthesis. It was shown that the fetal liver and the fetal adrenals contain a high level of 16 α -hydroxylase activity and much of the DS formed in the fetal adrenal glands was converted to 16 α -hydroxydehydroisoandrosterone sulfate (16 α -OHDS); and it was shown that this compound, in turn, was converted to E3 in the placenta (1, 35). Although extensive 16 α -hydroxylation of DS also occurs in the maternal compartment (23, 36), the rate of production of DS by the maternal adrenal cortex is much less than that by the fetal adrenal; as a consequence, approximately 90% of the E3 produced in human pregnancy is formed by the placental conversion of fetal plasma 16 α -OHDS to E3 (1). Thus the rate of placental E3 secretion is not solely a function of placental metabolism but rather represents the culmination of a complex interrelationship between the rate of fetal adrenal DS secretion, 16 α -OHDS formation, and placental aromatization of fetal plasma 16 α -OH-C₁₉-steroids.

The identification of the critical role of the fetus in supplying precursors for placental estrogen production provided additional impetus to measure maternal E3 levels. It was hoped that such measurements would serve as a guide to the optimal management of complicated pregnancies. E3 measurements have become a commonly used laboratory test requested by many obstetricians who seek to evaluate fetal well-being. Many investigators have provided data that attest to the fact that strikingly diminished E3 levels are found after fetal death *in utero* (37). Indeed many investigators are in agreement with the proposition that fetal jeopardy or death is reflected by the finding of declining levels of E3 in maternal plasma or urine (37-39). Unfortunately, however, there are many instances in which the presence or absence of fetal jeopardy is unquestioned and yet clearly E3 values do not accurately reflect fetal well-being, e.g. Rh-immunization with an affected fetus, certain types of diabetes mellitus, placental sulfatase deficiency, and abnormalities in maternal renal function.

The critical question is whether knowledge of the levels of maternal plasma or urinary E3 provides new information concerning the status of the fetus and whether such information is valuable in idealizing the management of a given high-risk pregnancy. Most importantly, the question is whether the use of this test (E3 levels) will bring about a decrease in perinatal mortality and/or morbidity. In the only prospective, controlled study known to have been conducted, the results supported the position that knowledge of the levels of estrogen in maternal plasma was of little or no utility in selecting management of the pregnancies of women whose fetuses were considered at high risk (40). Rather, this prospective study concluded that an increase in perinatal mortality may occur as a consequence of premature delivery of women with unexplained decreases or recurrent low estriol values but with no clinical findings to indicate that fetal jeopardy was present.

We are of the view that no measurement of steroid hormone(s) or metabolites during human pregnancy can yet provide the physician with additional information that will materially aid in the choice of management of high-risk pregnancies.

HUMAN CHORIONIC GONADOTROPIN

Human chorionic gonadotropin (hCG) is a glycoprotein that is likely produced by the syncytiotrophoblast of the placenta. Thiede & Choate (41) demonstrated, by immunofluorescent techniques, that hCG is concentrated in these cells. Two dissimilar subunits, designated α and β , comprise hCG. When these two subunits are separated, neither has intrinsic biologic activity (42), but after recombination approximately 66% of the native biologic

activity is regained. The α subunit of one glycoprotein hormone may be combined with the β subunit of a separate glycoprotein hormone, and such recombinations give rise to a molecule possessing biologic activity characteristic of the native hormone from which the β subunit was derived (43). Thus the α subunits of the various glycoprotein hormones are quite similar. Luteinizing hormone (LH) is immunologically and physiologically similar to hCG: Indeed, the β subunits as well as the α subunits of these two hormones are similar. However, the β subunit of hCG contains a richer content of proline and carbohydrate. Apparently this relatively small difference in amino acid and carbohydrate content makes it possible to develop antibodies highly specific for the β subunit of hCG, and such antibodies are utilized in radioimmunoassays for hCG.

Since the original demonstration by Aschheim & Zondek (44) of the pregnancy hormone in urine, the detection of hCG has been the basis of most pregnancy tests. During this time the methods for assaying hCG have varied. Earlier tests employed bioassay and immunoassay techniques, and, more recently, radioimmunoassay has become available. Bioassay procedures (being primarily of historic interest) required from 12 hours to as long as 5 days to complete. Various immunologic tests of hCG remain the most widely used pregnancy tests. Some difficulty may be encountered when utilizing immunoassay to measure hCG because of the similarity of LH and hCG, i.e. the antibodies utilized in most immunoassay tests recognize LH as well as hCG. Some believe that distinction between LH and hCG present in small amounts is helpful in the management of patients who have had hydatidiform moles or choriocarcinoma.

The rate of excretion of hCG in the urine during pregnancy gradually increases, attaining the highest levels between the 9th and 14th weeks of pregnancy. The levels of hCG found in urine and in serum are similar. After the 14th week of pregnancy, the levels of hCG in maternal serum and urine decline gradually, reaching a nadir at approximately 20 weeks of gestation. Significant variation in the level of hCG occurs among different pregnant women. Higher levels of hCG are sometimes found in women pregnant with multiple fetuses (45) and in women whose pregnancies are complicated by Rh-isoimmunization with fetal erythroblastosis (45, 46). Hydatidiform mole and choriocarcinoma also represent situations in which hCG levels are elevated (47).

One function of hCG is believed to be the maintenance of the corpus luteum during the first few weeks of pregnancy. Human chorionic gonadotropin may also serve to stimulate steroidogenesis within the fetal testes in early pregnancy, and a role for hCG in the immunologic privilege of the trophoblast has been proposed (48). Using radioimmunoassay procedures with antibodies highly specific for the β subunit of hCG, the presence of

immunoreactive hCG in maternal serum can be detected as early as eight days following ovulation in a fertile cycle (49). This technique allows early detection of pregnancy in women undergoing ovulation induction and/or artificial insemination. With such methods, it has been shown that the production rate of hCG is similar in pregnancies in which ovulation occurred spontaneously and in those in which ovulation was induced (50, 51). In several studies it was found that the renal clearance of hCG does not vary during the course of pregnancy. Thus, variations in serum and urine concentrations of hCG are a reflection of the changes in secretion rate and/or inactivation rather than altered renal excretion. The mechanism(s) that controls the rate of hCG production is unknown. In those trophoblastic diseases in which a fetus is not present, e.g. choriocarcinoma and/or hydatidiform mole, the levels of hCG are strikingly elevated. In other instances in which a fetus is present, e.g. toxemia of pregnancy (45), RH-isoimmunization (46), and maternal diabetes mellitus, hCG levels are sometimes increased. Interestingly, it has been observed that pregnancies with female fetuses are associated with higher maternal hCG levels than are those with male fetuses (52).

Knowledge of the levels of hCG is important in selecting optimal management of patients who have had hydatidiform moles and/or choriocarcinoma. It is well established that serum and urinary hCG levels are strikingly increased in women with hydatidiform mole pregnancies compared to those with normal pregnancies of similar gestational age. There are, however, overlaps in the levels of hCG in women with neoplastic trophoblastic disease with those found in normal pregnancies, but levels of hCG greater than 500 IU/ml of serum are virtually diagnostic of hydatidiform mole. The selection of clinical management of women with neoplastic trophoblastic disease has been facilitated by the development of radioimmunoassay procedures specific for the β subunit of hCG. Bioassay and hemagglutination inhibition assays are not considered reliable at levels below 750–3000 mIU/ml, and the lack of specificity of the antibodies used in such tests makes it impossible to distinguish between LH and hCG. Previously, attempts were made to compensate for the lack of specificity of the antibody used in such tests by administering oral contraceptive steroids to the test subject in an attempt to suppress pituitary LH secretion and thus avoid interference in the detection of hCG. Careful monitoring of women who have had molar pregnancies is essential, since approximately 50% of cases of choriocarcinoma develop following hydatidiform mole. Even though only 1–2% of all molar pregnancies are followed by choriocarcinoma, this previously fatal disease, if contained within the uterus, is now largely curable with appropriate chemotherapy. The measurement of hCG in attempts to evaluate the existence of ectopic pregnancies prior to the

advent of measurement of the β subunit of hCG were of little value. The finding of no hCG in serum seems to exclude the possibility of an ectopic pregnancy (51). While the serum levels of hCG are generally lower in women with ectopic pregnancies than in normal pregnant women, differentiation between ectopic pregnancies and abnormal intrauterine pregnancies on the basis of serum hCG levels alone does not seem possible at this time.

HUMAN PLACENTAL LACTOGEN

Human placental lactogen (hPL) was first described by Ito & Higashi (53) and isolated by Josimovich & McClaren in 1962 (54). It has both lactogenic and growth-hormone-like properties, and for this reason some investigators refer to this placental hormone as chorionic somatomammotropin. The pattern of secretion of hPL in human pregnancy is parallel to placental weight and fetal weight: hPL can be detected in the serum of pregnant women very early in pregnancy, and the levels rise steadily throughout most of pregnancy reaching a plateau at approximately 36–37 weeks when levels are approximately 5–8 $\mu\text{g/ml}$ (55). In pregnancies with multiple fetuses, levels of hPL are higher than those of normal singleton pregnancies (56). Levels of hPL as high as 40 $\mu\text{g/ml}$ have been found in quadruplet and quintuplet pregnancies (57). The hPL levels in the fetal circulation are low throughout gestation, somewhat higher in amniotic fluid, but the highest levels are found in the maternal circulation. Following delivery of the placenta, there is a rapid disappearance of plasma hPL; within 24 hours after delivery, plasma hPL is undetectable. The metabolic clearance rate (MCR) of hPL is approximately 175 liters/day, and the MCR does not appear to be affected by its plasma concentration (58). The production rate of hPL in late human pregnancy is approximately 1 g/day. Thus, the turnover rate of hPL is greater than that described for any other polypeptide hormone. The physiologic role of hPL during human pregnancy is poorly understood. However, hPL appears to be a hormone secreted by fetal tissue that has major metabolic effects on the maternal organism, some of these effects may participate in ensuring the nutritional needs of the fetus (62). The rate of secretion of hPL in late pregnancy seems to be minimally affected by maternal hypoglycemia, hyperglycemia, hyperaminoacidemia, or stress. However, during maternal starvation a marked rise in serum hPL is observed (59). During normal human pregnancy blood glucose levels are decreased, insulin secretion is increased, and there develops an increased resistance to insulin and an elevation of maternal plasma free fatty acids. These metabolic events have been attributed to the actions of hPL, actions that lead to impaired glucose uptake directly and/or indirectly by stimula-

tion of free fatty acid release. Thus hPL participates in a number of metabolic events, which include lipolysis, elevation of serum free fatty acids, and inhibition of glucose uptake and gluconeogenesis in the maternal organism.

Constantly in search of means of evaluating fetal well-being in human pregnancy, many investigators have measured maternal plasma hPL levels and have sought to ascertain if such values are useful in evaluating high-risk obstetric patients. One major obstacle in interpreting the results of such measurements is the wide range of normal values, e.g. normal levels of hPL of 3.3 $\mu\text{g/ml}$ were reported by Samaan et al (60), whereas values of 25 $\mu\text{g/ml}$ were reported by Beck et al (61). This variability no doubt represents methodologic differences. Presently most investigators have chosen the value of 4 $\mu\text{g/ml}$ as the lower limit of normal for pregnancies of more than 30 weeks gestation. In a study of 204 normal women, Spellacy et al (62) found that only 1% of values were less than 4 $\mu\text{g/ml}$ after 30 weeks of gestation. Levels of hPL correlate well with placental as well as fetal weight (63). There are, however, conflicting reports regarding hPL values in pregnant women who are beyond term. Some investigators have found low values in postterm patients while others found normal values. Levels of hPL have also been measured by several investigators who sought to use such values in the management of women with threatened abortions. Specifically, they asked if hPL values could be used to predict accurately which women were most likely to abort. Gartside & Tindall (64) found that based on serum hPL levels they could correctly predict pregnancy outcome in 86% of cases between 9 and 19 weeks of gestation, while an incorrect prediction was made in only 3%. In a study of 236 patients with uterine bleeding in early pregnancy, Niven et al (65) found that after the ninth week of pregnancy all women who subsequently aborted had hPL values below normal; conversely, the 83 patients whose pregnancies continued had hPL values within the normal range.

The hPL values are of no clinical value in selecting optimal management of pregnancies complicated by Rh-isoimmunization. In most pregnant women with diabetes mellitus, generally higher levels of hPL are found and most investigators have concluded that low levels of hPL in such women are indicative of placental insufficiency (66). In a study of women with pregnancy-induced hypertension, Spellacy et al (67) found that if the hPL values were in the fetal danger zone, i.e. less than 4 ng/ml, there was a still-birth rate of 24%. Spellacy et al also studied a group of women in whom 4025 hPL values were obtained in 949 pregnant women with hypertension (67). These women were categorized on the basis of time of appearance of hypertension, the degree of hypertension, and the presence or

absence of proteinuria, and the parity of the mother. They found an increase in the number of hPL values that were in the fetal danger zone in such women and a higher incidence of low values if hypertension was present both antepartum and during labor. There was, however, no correlation between the appearance of proteinuria and the hPL values. They concluded that lower levels of hPL were present in women with chronic hypertensive vascular disease, presumably reflecting placental malfunction in such pregnancies. These results were confirmed by Letchworth & Chard (68), who found generally lower levels of hPL in gravidas with hypertension compared to those found in normal pregnant women. Lindberg & Nilsson (69) studied 98 women with preeclampsia. They found that in the 47 patients whose hPL values were less than $4 \mu\text{g/ml}$, the perinatal death rate was 13%, and intrauterine growth retardation was present in 57%. The results of many studies are difficult to evaluate, however, since the nature and severity of the hypertension were not defined. Nonetheless most investigators who have studied large groups of women tend to agree that the levels of hPL are reduced in pregnancies complicated by preeclampsia and hypertension, and that the lowest values are found in the women with the most severe hypertension.

According to Spellacy, Buhi & Birk (70), who reported a prospective random study involving 2733 women in whom serial hPL measurements were performed at each prenatal visit, estimation of hPL levels is of value in predicting fetal well-being. The frequency of hPL values less than $4 \mu\text{g/ml}$ was similar in each group (8.6% in the treatment group as compared to 8.2% in the control group). Neonatal mortality rates were similar in each group; however, the patients in whom the hPL values were reported to the obstetricians had a significantly lower fetal death rate (2.6 vs 14.2%) and thus lower perinatal mortality rate (3.4 vs 15%) when compared to the group in whom the hPL values were not reported to the obstetric team. They conclude that the use of routine prenatal hPL measurements is useful in the management of high-risk pregnancies. Others have questioned this conclusion (40). We are of the view that the measurement of hPL levels in high-risk pregnancy is of little utility in selecting management of such pregnancies and provides little or no new information not available from clinical assessment alone.

HUMAN CHORIONIC THYROTROPIN

A substance extracted from normal human placenta that exhibited thyrotropic activity was first reported by Hennen (71) in 1965. The name human chorionic thyrotropin (hCT) was proposed by Li (72) and readily accepted in 1968. The possibility of a thyroid stimulator present in the placenta

during pregnancy raises many questions. It has long been appreciated that several parameters of thyroid function are altered during pregnancy. However, normal pregnancy is not accompanied by clinical evidence of hyperthyroidism.

Thyroid enlargement is common during pregnancy and increased levels of serum protein-bound iodine (PBI) are also present during normal pregnancy. The elevated PBI is largely a phenomenon secondary to a rise in serum thyroxin-binding globulin, which is the consequence of elevated secretion of estrogen during pregnancy. There also appears to be an increase in thyroxin-binding capacity. However, the absolute concentration of free thyroxin remains constant throughout pregnancy. Human chorionic thyrotropin is a glycoprotein similar to pituitary thyrotropin; however, it is less anionic because it contains less sialic acid than does pituitary thyrotropin. Of all thyroid-stimulating hormones currently characterized, hCT appears to be more nearly like bovine thyrotropin than human pituitary thyrotropin (73). The physiologic role of hCT during pregnancy is at best unclear. There are, however, several reported cases of hyperthyroidism associated with both hydatidiform moles and other trophoblastic tumors, including hyperthyroidism in two men with metastatic choriocarcinoma. However, in such subjects hCG levels are strikingly increased and hCG has been shown to be a thyrotropin (74, 75).

Literature Cited

1. Siiteri, P. K., MacDonald, P. C. 1966. Placental estrogen biosynthesis during human pregnancy. *J. Clin. Endocrinol. Metab.* 26:751-61
2. Frandsen, V. A., Stakeman, G. 1963. The urinary excretion of oestriol during the early months of pregnancy. Normal and pathological cases. *Acta Endocrinol.* 44:196-208
3. Pearlman, W. H. 1957. [16-³H]Progesterone metabolism in advanced pregnancy and in oophorectomized-hysterectomized women. *Biochem. J.* 67:1-5
4. Siler-Khodr, T. M., Khodr, G. S. 1978. Luteinizing hormone releasing factor content of the human placenta. *Am. J. Obstet. Gynecol.* 130:216-19
5. Everett, R. B., Worley, R. J., MacDonald, P. C., Gant, N. F. 1978. Modification of vascular responsiveness to angiotensin II in pregnant women by intravenously infused 5 α -dihydroprogesterone. *Am. J. Obstet. Gynecol.* 132: 359-62
6. Davies, I. J., Ryan, K. J. 1972. Comparative endocrinology of gestation. *Vitam. Horm.* 30:223-79
7. Cassmer, O. 1959. Hormone production of the isolated human placenta. *Acta Endocrinol. (Copenhagen): Suppl.* 45, p. 1
8. Hellig, H., Gattereau, D., Lefebvre, Y., Bolté, E. 1970. Steroid production from plasma cholesterol. I. Conversion of plasma cholesterol to placental progesterone in humans. *J. Clin. Endocrinol. Metab.* 30:624-31
9. Simpson, E. R., Porter, J. C., Milewich, L., Bilheimer, D. W., MacDonald, P. C. 1978. Regulation by plasma lipoproteins of progesterone biosynthesis and 3-hydroxy-3-methyl glutaryl coenzyme A reductase activity in cultured human choriocarcinoma cells. *J. Clin. Endocrinol. Metab.* In press
10. Simpson, E. R., Bilheimer, D. W., MacDonald, P. C., Porter, J. C. 1978. Uptake and degradation of plasma lipoproteins by human choriocarcinoma cells in culture. *Endocrinology.* In press
11. Goldstein, J. L., M. S. Brown 1974. Binding and degradation of low density lipoproteins by cultured human fibroblasts. *J. Biol. Chem.* 249:5153-62

12. Escarcena, L., Clark, H., Gursipide, E. 1978. Contribution of maternal circulation to blood-bovine progesterone in the fetus. 1. Studies on human subjects. *Am. J. Obstet. Gynecol.* 130:452-65
13. Harbert, G. M., McGaughey, H. S. Jr., Scoggin, W. A., Thornton, W. N. 1964. Concentration of progesterone in newborn and maternal circulations at delivery. *Obstet. Gynecol.* 23:413-26
14. Solomon, S., Fuchs, F. 1971. Progesterone and related neutral steroids. In *Endocrinology of Pregnancy*, ed. F. Fuchs, A. Kloppe, pp. 66-100. New York: Harper & Row. 359 pp. 1st ed.
15. Ryan, K. J. 1973. Steroid hormones in mammalian pregnancy. In *Handbook of Physiology*, ed. R. O. Greep, E. B. Astwood, pp. 285-93. Am. Physiol. Soc. Baltimore, Md: Williams & Wilkins
16. Milewich, L., Gomez-Sanchez, C. E., Madden, J. D., MacDonald, P. C. 1975. Isolation and characterization of 5 α -pregnane-3,20-dione and progesterone in peripheral blood of pregnant women. Measurement throughout pregnancy. *Gynecol. Invest.* 6:291-306
17. MacDonald, P. C., Porter, J. C., Schwarz, B. E., Johnston, J. M. 1978. Initiation of parturition in the human female. *Semin. Perinatol.* 2:273-86
18. Frandsen, V. A., Stakeman, G. 1961. The site of production of oestrogenic hormones in human pregnancy. Hormone excretion in pregnancy with anencephalic foetus. *Acta Endocrinol. (Copenhagen)* 38:383-91
19. MacDonald, P. C., Siiteri, P. K. 1964. Study of estrogen production in women with hydatidiform mole. *J. Clin. Endocrinol. Metab.* 24:685-90
20. Siiteri, P. K., MacDonald, P. C. 1963. The utilization of circulating dehydroisoandrosterone sulfate for estrogen synthesis during human pregnancy. *Steroids* 2:713-30
21. MacDonald, P. C., Siiteri, P. K. 1965. Origin of estrogen in women pregnant with an anencephalic fetus. *J. Clin. Invest.* 44:465-74
22. Diczfalusy, E., Mancuso, S. 1969. *Fetus and Placenta*, p. 191. Oxford: Blackwell Scientific. 629 pp.
23. Madden, J. D., Siiteri, P. K., MacDonald, P. C., Gant, N. F. 1976. The pattern and rates of metabolism of maternal plasma dehydroisoandrosterone sulfate in human pregnancy. *Am. J. Obstet. Gynecol.* 125:915-20
24. Tabei, T., Heinrichs, W. L. 1976. Diagnosis of placental sulfatase deficiency. *Am. J. Obstet. Gynecol.* 124:409-14
25. Gant, N. F., Madden, J. D., Siiteri, P. K., MacDonald, P. C. 1972. A sequential study of the metabolism of dehydroisoandrosterone sulfate in primigravid pregnancy. *Proc. Int. Congr. Endocrinol., 4th, Washington*, pp. 1026-31
26. Gant, N. F., Hutchinson, H. T., Siiteri, P. K., MacDonald, P. C. 1971. Study of the metabolic clearance rate of dehydroisoandrosterone sulfate in pregnancy. *Am. J. Obstet. Gynecol.* 111: 555-61
27. Gant, N. F., Madden, J. D., Chand, S., Worley, R. J., Strong, J. D., MacDonald, P. C. 1976. Metabolic clearance rate of dehydroisoandrosterone sulfate. V. Studies of essential hypertension complicating pregnancy. *Obstet. Gynecol.* 47:319-26
28. Gant, N. F., Madden, J. D., Chand, S., Worley, R. J., Siiteri, P. K., MacDonald, P. C. 1976. The metabolic clearance rate of dehydroisoandrosterone sulfate: VI. Studies of eclampsia. *Obstet. Gynecol.* 47:327-30
29. Gant, N. F., Madden, J. D., Siiteri, P. K., MacDonald, P. C. 1975. The metabolic clearance rate of dehydroisoandrosterone sulfate. III. The effect of thiazide diuretics in normal and future pre-eclamptic pregnancies. *Am. J. Obstet. Gynecol.* 123:159-63
30. Gant, N. F., Madden, J. D., Siiteri, P. K., MacDonald, P. C. 1976. The metabolic clearance rate of dehydroisoandrosterone sulfate. IV. Acute effects of induced hypertension, hypotension, and naturesis in normal and hypertensive pregnancies. *Am. J. Obstet. Gynecol.* 124:143-48
31. Worley, R. J., Everett, R. B., MacDonald, P. C., Gant, N. F. The placental clearance of maternal plasma dehydroisoandrosterone sulfate through estradiol. II. A comparison of normal singleton and twin pregnancies. *Am. J. Obstet. Gynecol.* Submitted
32. Worley, R. J., Everett, R. B., MacDonald, P. C., Gant, N. F. 1975. Placental clearance of dehydroisoandrosterone sulfate and pregnancy outcome in three categories of hospitalized patients with pregnancy-induced hypertension. *Gynecol. Invest.* 6:28 (Abstr.)
33. Shoemaker, E. S., Gant, N. F., Madden, J. D., MacDonald, P. C. 1973. The effect of thiazide diuretics on placental function. *Tex. Med.* 69:109-15
34. Ryan, K. J. 1959. Metabolism of C-16-oxygenated steroids by human placenta: The formation of estriol. *J. Biol. Chem.* 234:2006-8

35. Bolté, E., Wiqvist, N., Diczfalussy, E. 1966. Metabolism of dehydroepiandrosterone and dehydroepiandrosterone sulfate by the human foetus at midpregnancy. *Acta Endocrinol.* 52:583-97
36. Madden, J. D., Gant, N. F., MacDonald, P. C. 1978. Study of the kinetics of conversion of maternal plasma dehydroisoandrosterone sulfate to 16 α -hydroxydehydroisoandrosterone sulfate, estradiol, and estriol. *Am. J. Obstet. Gynecol.* 132:392-95
37. Frandsen, V., Jorgensen, P. I., Svenstrup, B. 1970. The adventures of one thousand and one obstetrical patients assessed by oestriol excretion. *Ann. Clin. Res.* 2:354-64
38. Greene, J. W., Touchstone, J. C. 1963. Urinary estriol as an index of placental function. *Am. J. Obstet. Gynecol.* 85:1-9
39. Frandsen, V. A., Stakeman, G. 1963. The clinical significance of oestriol estimations in late pregnancy. *Acta Endocrinol.* 44:183-95
40. Duenhoelter, J. H., Walley, P. J., MacDonald, P. C. 1976. An analysis of the utility of plasma immunoreactive estrogen measurements in determining delivery time of gravidas with a fetus considered at high risk. *Am. J. Obstet. Gynecol.* 125:889-98
41. Theide, H. A., Choate, J. W. 1963. Chorionic gonadotropin localization in the human placenta by immunofluorescent staining. II. Demonstration of HCG in the trophoblast and amnion epithelium of immature and mature placentas. *Obstet. Gynecol.* 22:433-43
42. Morgan, F. J., Canfield, R. E. 1971. Nature of the subunits of human chorionic gonadotropin. *Endocrinology* 88: 1045-53
43. Pierce, J. G., Bahl, O. P., Cornwell, J. S., Swaminathan, N. 1971. Biologically active hormones prepared by recombination of the α chain of human chorionic gonadotropin and the hormone-specific chain of bovine thyrotropin or of bovine luteinizing hormone. *J. Biol. Chem.* 246:2321-24
44. Aschheim, S., Zondek, B. 1927. Anterior pituitary hormone and ovarian hormone in the urine of pregnant women. *Klin. Wochenschr.* 6:1322-28
45. Crosignani, P. G., Trojsi, L., Attanasio, A. E. M., Lombroso Finzi, G. S. 1974. Value of HCG and HCS measurement in clinical practice. *Obstet. Gynecol.* 44:673-81
46. Bradbury, J. T., Goplerud, C. P. 1963. Serum chorionic gonadotropin studies in sensitized Rh-negative patients. *Obstet. Gynecol.* 21:330-33
47. Yen, S. S. C., Pearson, O. H., Rankin, J. S. 1968. Radioimmunoassay of serum chorionic gonadotropin and placental lactogen in trophoblastic disease. *Obstet. Gynecol.* 32:86-93
48. Teasdale, F., Adcock, E. W., August, C. S., Cox, S., Battaglia, F. C., Naughton, M. A. 1973. Human chorionic gonadotropin: Inhibitory effect on mixed lymphocyte cultures. *Gynecol. Invest.* 4:263-69
49. Van Leusden, H. A. 1976. *Plasma Hormone Assays in Evaluation of Fetal Well-Being*, ed. A. Kloppe, pp. 48-71. New York: Churchill Livingstone. 177 pp.
50. Mishell, D. R., Nakamura, R. M., Barberia, J. M., Thorneycroft, I. H. 1974. Initial detection of human chorionic gonadotropin in serum in normal human gestation. *Am. J. Obstet. Gynecol.* 118:990-94
51. Kosasa, T. S., Levesque, L. A., Goldstein, D. P., Taymor, M. L. 1974. Clinical use of a solid-phase radioimmunoassay specific for human chorionic gonadotropin. *Am. J. Obstet. Gynecol.* 119:784-91
52. Boroditsky, R. S., Reyes, F. I., Winter, J. S. D., Faiman, C. 1974. Serum human chorionic gonadotropin and progesterone patterns in the last trimester of pregnancy: Relationship to fetal sex. *Am. J. Obstet. Gynecol.* 121:238-41
53. Ito, Y., Higashi, K. 1961. Studies on the prolactin-like substance in human placenta II. *Endocrinol. Jpn.* 8:279-87
54. Josimovich, J. B., MacLaren, J. A. 1962. Presence in the human placenta and term serum of a highly lactogenic substance immunologically related to pituitary growth hormone. *Endocrinology* 71:209-20
55. Grumbach, M. M., Kaplan, S. L., Sciarra, J. J., Burr, I. M. 1968. Chorionic growth hormone-prolactin (CGP): Secretion, disposition, biologic activity in man, and postulated function as the "growth hormone" of the second half of pregnancy. *Ann. NY Acad. Sci.* 148:501-31
56. Josimovich, J. B., Kosor, B., Boccella, L., Mintz, D. H., Hutchinson, D. L. 1970. Placental lactogen in maternal serum as an index of fetal health. *Obstet. Gynecol.* 36:244-250
57. Josimovich, J. B., Levitt, M. J., Grumbach, M. M., Kaplan, S. L., Vinik, A. 1973. Human chorionic somatomammotropin (HCS). In *Methods in Investi-*

- gative and Diagnostic Endocrinology*, ed. S. A. Berson, R. S. Yalow, pp. 787-819. Amsterdam: North-Holland
58. Kaplan, S. L., Gursipde, E., Sciarra, J. J., Grumbach, M. M. 1968. Metabolic clearance rate and production rate of chorionic growth hormone-prolactin in late pregnancy. *J. Clin. Endocrinol. Metab.* 28:1450-60
 59. Kim, Y. J., Felig, P. 1971. Plasma chorionic somatomammotropin levels during starvation in mid-pregnancy. *J. Clin. Endocrinol. Metab.* 32:864-67
 60. Samaan, N., Yen, S. C. C., Friesen, H., Pearson, O. H. 1966. Serum placental lactogen levels during pregnancy and in trophoblastic disease. *J. Clin. Endocrinol. Metab.* 26:1303-8
 61. Beck, P., Parker, M. L., Daughaday, W. H. 1965. Radioimmunologic measurement of human placental lactogen in plasma by a double antibody method during normal and diabetic pregnancies. *J. Clin. Endocrinol. Metab.* 25: 1457-62
 62. Spellacy, W. N., Teoh, E. S., Buhi, W. C. 1970. Human chorionic somatomammotropin (HCS) levels prior to fetal death in high-risk pregnancies. *Obstet. Gynecol.* 35:685-89
 63. Spellacy, W. N., Teoh, E. S., Buhi, W. C., Birk, S. A., McCreary, S. A. 1971. Value of human chorionic somatomammotropin in managing high-risk pregnancies. *Am. J. Obstet. Gynecol.* 109: 588-98
 64. Gartside, M. W., Tindall, V. R. 1975. The prognostic value of human placental lactogen (HPL) levels in threatened abortion. *Br. J. Obstet. Gynaecol.* 82: 303-9
 65. Niven, P. A. R., Landon, J., Chard, T. 1972. Placental lactogen levels as guide to outcome of threatened abortion. *Br. Med. J.* 3:799-801
 66. Selenkow, H. A., Varma, K., Younger, D., White, P., Emerson, K. 1971. Patterns of serum immunoreactive human placental lactogen (IR-HPL) and chorionic gonadotropin (IR-HCG) in diabetic pregnancy. *Diabetes* 20:696-706
 67. Spellacy, W. N., Buhi, W. C., Birk, S. A., McCreary, S. A. 1974. Distribution of human placental lactogen in the last half of normal and complicated pregnancies. *Am. J. Obstet. Gynecol.* 120: 214-23
 68. Letchworth, A. T., Chard, T. 1972. Human placental lactogen levels in pre-eclampsia. *J. Obstet. Gynaecol. Br. Commonw.* 79:680-83
 69. Lindberg, B. S., Nilsson, B. A. 1973. Human placental lactogen (HPL) levels in abnormal pregnancies. *J. Obstet. Gynaecol. Br. Commonw.* 80:1046-53
 70. Spellacy, W. N., Buhi, W. C., Birk, S. A. 1975. The effectiveness of human placental lactogen measurements as an adjunct in decreasing perinatal deaths. Results of a retrospective and a randomized controlled prospective study. *Am. J. Obstet. Gynecol.* 121:835-44
 71. Hennen, G. P. 1965. Detection and study of a human-chorionic-thyroid-stimulating factor. *Arch. Int. Physiol. Biochim.* 73:689-95
 72. Li, C. H. 1969. In *Protein and Polypeptide Hormone*, ed. M. Margoulies. Amsterdam: Excerpta Med. Int. Congr. Ser., No. 161, p. 797
 73. Hennen, G. P., Pierce, J. G., Freychet, P. 1969. Human chorionic thyrotropin: Further characterization and study of its secretion during pregnancy. *J. Clin. Endocrinol. Metab.* 29:581-94
 74. Kenimer, J. G., Hershman, J. M., Higgins, H. P. 1975. The thyrotropin in hydatidiform moles in human chorionic gonadotropin. *J. Clin. Endocrinol. Metab.* 40:482-91
 75. Silverberg, J., O'Donnell, J., Sugenoja, A., Row, V. V., Volpe, R. 1978. Effect of chorionic gonadotropin on human thyroid tissue *in vitro*. *J. Clin. Endocrinol. Metab.* 46:420-24

THE EFFICACY OF BIOFEEDBACK THERAPY

♦7337

Martin T. Orne, M.D., Ph.D.

The Institute of Pennsylvania Hospital and University of Pennsylvania,
Philadelphia, Pennsylvania 19139

WHAT IS BIOFEEDBACK?

The effective functioning of any system—electrical, physiological, even social—is contingent upon the return of information to the system concerning its performance. The loop that provides the system with information about its output is called feedback. All of the homeostatic mechanisms of the body depend upon complicated interrelated feedback mechanisms that serve to stabilize the internal environment. Similarly, all skill learning depends upon knowledge of results. Practice makes perfect only if the individual is aware of how well he performs.

An illustration of the significance of feedback involves the acquisition of speech. The child learns to speak by imitating the sounds he hears, and the adequacy of learning depends upon his ability to compare the sound of the words he generates with that of the words generated by those around him. For this reason, individuals who become deaf later in life continue to speak without much difficulty, but children born deaf were, until relatively recently, unable to speak. Only when special techniques were developed by which the deaf person could learn the consequences of his attempts to speak, using visual and tactile cues to make up for his hearing deficit, did it become possible for totally deaf individuals to learn intelligible speech.

In a sense, teaching a deaf person to speak is an example of biofeedback. It involves providing alternative feedback channels to compensate for those not available to the individual so that he may acquire volitional control over a specific bodily function. It is therefore not surprising that the earliest applications of biofeedback were in the context of rehabilitation. These included not only teaching the deaf to speak or the blind to “see” with their

hands and ears, but also retraining of muscles after injury or a paralytic disease, particularly when it was necessary for the individual to utilize the remaining muscle groups in a novel manner. Though the accomplishments of biofeedback in rehabilitation were impressive, the full implications of providing new channels of feedback to the individual were not generally appreciated because they were used to teach "ordinary" skills that are within the behavioral repertoire of most individuals.

BASIS FOR THE DEVELOPMENT OF BIOFEEDBACK THERAPY

The exciting aspect of the biofeedback technology that began to evolve during the 1960s is that it seems to provide methods to teach volitional control over physiological functions which were generally believed to be involuntary. Most of the research relevant to this work was published in psychological journals and is in large part couched in terms of a major theoretical controversy concerning the nature of visceral learning.

For many years it was accepted that physiological changes such as increases in heart rate, blood pressure, electrodermal response, alterations of brain waves, and the like could only be brought about in a reflex fashion. Classical Pavlovian conditioning, where the organism is essentially passive, automatically responding to what is being done to it, was seen as the mechanism of visceral learning. For example, a painful shock may elicit an increase in heart rate. If the shock is consistently preceded by a tone, the tone will begin to elicit the same response in a lawful, predictable fashion. The issue is not one of motivation; rather, the response is evoked involuntarily in a reflex-like manner, initially by the shock and subsequently by the tone as well. By using such procedures, any physiological change that could consistently be elicited involuntarily by a particular stimulus as an unconditional reflex could also be conditioned so that it would be evoked by different stimuli as a conditional reflex. It was assumed that "thoughts" could serve as conditional stimuli, and it was thereby explained how man could learn a modicum of volitional control over involuntary physiological functions.

An entirely different mechanism, however, was deemed relevant to the learning of volitional behavior. This process, known as instrumental or operant conditioning, depends upon reinforcing the organism when it is producing a desired response. In animal research, for example, the rate and consistency of lever pressing is easily controlled by the frequency and consistency with which the hungry animal is contingently reinforced (e.g. rewarded with food) when he makes the desired response. Whereas in Pavlovian conditioning the organism's "response" involves a change nor-

mally elicited involuntarily by a reflex, in operant or instrumental conditioning the response is any item of behavior within the organism's normal repertoire. In this sense, a reinforcer in this type of conditioning merely serves to make it worthwhile for the organism to do something that is already within its capabilities.

It seems somewhat strange, however, to request an individual to raise or lower his heart rate, since this would seem to be outside of his capabilities regardless of the rewards that might be offered. Thus, a two-factor theory of learning, classical conditioning for visceral learning and instrumental conditioning for the acquisition of skills and volitional behavior, originally outlined by Skinner (1, also see 2), was widely accepted.

Early Biofeedback Studies in Man

During the 1960s a number of investigators began to apply the operant model as a means of modifying visceral activity; in other words, negative and positive consequences contingent upon an apparently involuntary visceral response were used to modify that response. For example, Engel (3) initiated studies to control some cardiac arrhythmias. Noting that slight changes in heart rate would virtually eliminate the arrhythmias, he set about training patients to control their heart rate. Electronic equipment was used to compare the frequency on a beat-by-beat basis, and when subjects raised their heart rate even by a slight amount, a signal light was turned on. Engel then instructed subjects to keep the signal light on. Conceptualizing the light as a reward, he observed that subjects could indeed learn to increase their heart rate. Though the extent of control was limited, there was no doubt that subjects did indeed learn to alter their heart rate, and in some cases gained control over their arrhythmias. Control of heart rate in normal subjects was also documented by Shapiro et al (4), who went on to show that the change in heart rate was independent of changes in blood pressure. Kimmel (5), and subsequently Shapiro et al (6), documented operant control of the electrodermal response. Basmajian (7) showed that the firing of single motor neurons could similarly be brought under volitional control, a control sufficiently delicate to allow some subjects even to send Morse code in this fashion.

These studies, which were carried out largely with human subjects, seemed to show that, contrary to previous beliefs, appropriate technology could help individuals acquire volitional control over a number of different autonomic functions. Further, this control seemed to involve considerable specificity and could not be explained simply as the result of "thinking arousing thoughts." Nonetheless, there was considerable controversy concerning these studies, and a number of scientists were unwilling to accept the observations as compelling evidence that cortical control over auto-

nostic activity had been demonstrated. Rather, they argued that subjects merely learned to change their physiological response either by playing mental games (taking advantage of classically acquired conditional reflexes with their thoughts) or perhaps utilizing covert muscular activity. It remained for Miller and his associates to document the radically new nature of the emerging technology.

Biofeedback Research with Animals

In a now classic paper Miller (8) reported a number of studies carried out on animals who were curarized—thereby eliminating the possibility of mediation by skeletal muscle activity—which documented that a variety of visceral responses could be brought under operant control. For example, heart rate and blood pressure could be varied independently of each other, intestinal motility could be altered, and the specificity of the responses was sufficiently great that it was possible to train rats to increase blood flow in one ear while simultaneously decreasing it in the other ear (9). It was shown that operant procedures could produce profound changes of medical relevance when 7 of 43 curarized rats trained to slow their heart rate died from apparent heart failure while none of the 41 rats trained to increase their heart rate showed such an effect (10). This experiment not only dramatized the power of this new procedure but also highlighted why some physiologically oriented scientists had difficulty from a teleological point of view in conceiving that nature would entrust such an important function as heart rate to the volitional control of any organism. Nonetheless, the work of Miller and his associates effectively provided the scientific legitimacy needed for the emerging field. It appeared to resolve the lingering doubts about the possibility of achieving instrumental control over physiological processes and seemed to justify a major effort to develop this new technology into a useful therapeutic modality.

Electroencephalogram Biofeedback

Another major area of interest developed independently and involved the work of Kamiya (11), who appeared to demonstrate that brain waves, most notably alpha wave activity, could be brought under operant control. Though brain wave activity had not really been considered analogous to visceral activity, it had nonetheless been assumed to be beyond volitional control. Kamiya not only seemed to document that subjects increased alpha density in a dimly lit environment but found that these increases appeared to persist beyond the training. His subjects reported that they preferred the experience of increasing alpha density to that of blocking alpha. Finally, he found that many subjects reported feelings that had traditionally been associated with meditative disciplines, such as calm, passive, relaxed, pleas-

ant, and the like, while producing large amounts of high density alpha in their EEG.¹

The findings of Kamiya were widely reported in the media and supported by the work of Brown (12), Nowlis & Kamiya (13), Hart (14), and others. Though these studies tended to lack systematic controls, they nonetheless caught the imagination of many serious scientists as well as the media, a considerable number of technologically oriented individuals, and a sizable segment of the counterculture.

Other work in EEG control showed that individuals could be trained to control their evoked cortical response (15), to vary the amount of theta in their EEG (16), and even to gain control over the sensory motor rhythms that increased their threshold for convulsions, making this type of training a potentially useful treatment for epilepsy (18).

BIOFEEDBACK AND RELAXATION

As has already been noted, the use of muscle tension feedback has a considerable history in rehabilitation. The application of this technique to lower the overall level of muscle tension did not develop, however, until the late 1960s. Stoyva & Budzynski (19) reported that frontalis muscle feedback could be employed to train people in general relaxation, and that while electromyogram (EMG) feedback was initially specific in its effects, it soon generalized so that as the tonic level of frontalis muscle tension was reduced, individuals became progressively more relaxed, indicated by the level of muscle tension in other muscle groups as well as in their verbal reports. This type of training has been widely applied in the treatment of tension headaches and has been used in the treatment of other pathology involving high arousal.

Another parameter used for biofeedback training is finger temperature. It has long been recognized that anxiety leads to peripheral vasoconstriction—the cold hand of the frightened individual. Objective psychophysiological studies have shown that finger temperature and forehead temperature vary inversely so that increasing finger temperature and decreasing forehead temperature are associated with increased relaxation and decreased anxiety. In the context of an early feedback study, Sargent et al (20) noted one of their patients who happened to have frequent migraine attacks became symptom-free in the context of an experiment involving training to raise finger temperature. This encouraging finding has led to the use of

¹Independently of Kamiya, Mulholland (17) reported that subjects could learn to increase or decrease alpha density. However, he did not generally observe the association between alpha density and subjective states reported by Kamiya.

finger temperature training as a treatment of migraine, based in part on the logic that the same process that causes vasoconstriction of the forehead is likely to decrease the vasodilation of the cerebral vessels, which is the generally accepted mechanism underlying acute migraine attacks, with highly encouraging results.

As was noted earlier, alpha feedback training was similarly reported as leading to a relaxed, comfortable, anxiety-free state. Thus, mental and physical relaxation were seen as important consequences of several kinds of biofeedback training. The importance of relaxation in the clinical application of biofeedback led to a renewed interest in relaxation training without biofeedback. The history of such training as part of various meditative disciplines goes back to early antiquity. The more systematic and scientific study of relaxation training began with the work of Schultz (21) in Germany and Jacobson (22) in the United States. Schultz developed a procedure he referred to as autogenic training (23), which he saw as embodying the major healing aspects of hypnosis stripped of some of its mystical overtones. Autogenic training involved a series of learned exercises to induce calmness and relaxation. They included heaviness of the hand and, most interestingly, warmth of the hand, followed by training involving coolness of the forehead, evenness of respiration, and so on. This procedure, requiring a considerable period of training and extensive practice on the part of the student, was seen as producing predictable neurophysiological changes with profound salutary effects on a wide range of medical disorders. Though almost unknown in the United States until the 1960s, this procedure has long been widely used in the German-speaking countries and in Japan. There is an extensive clinical literature on the application of this technique (24).

The work of Jacobson (22) on the use of progressive relaxation stems from a somewhat different theoretical framework and focuses almost exclusively on learned muscle relaxation. While the procedures are somewhat different, the end result is to train individuals to be able to relax profoundly. Some of the salutary consequences ascribed to this technique show considerable overlap with those claimed for autogenic training.

Other procedures that have recently aroused modest scientific and considerable popular interest, such as transcendental meditation (TM), seem to produce similar kinds of physical and mental relaxation. Indeed, in their popular book, Benson & Klipper (25) discuss the relaxation response as a generic concept which has profound therapeutic effects for the individual.

CURRENT STATUS OF BIOFEEDBACK RESEARCH AND THERAPY

Since biofeedback therapy involves both a novel approach to pathology and an effective technology, it has been applied to a wide range of medical

problems. The historical context that led to the introduction of biofeedback therapy is relevant because without the excitement generated among the public as well as within the profession the medical potential of biofeedback therapy would never have been realized. Fortunately, the clinical findings can now be evaluated on their merits, and the section below summarizes the results with some of the more important applications—results that, in some areas at least, are promising indeed.

It would be misleading, however, in an overview of this kind not to point out that the important studies with curarized animals that provided the scientific justification for much of the clinical research, while initially replicated, cannot now be reproduced. Though there is no difficulty in demonstrating statistically significant changes in visceral functions as a result of instrumental conditioning in curarized animals—leaving no doubt about the phenomenon—obtaining effects sufficiently large to be clinically significant eludes the present techniques (26). Similarly, alpha feedback training, which inspired much of the initial interest in biofeedback, has not thus far lived up to its initial promise (27). Finally, the techniques of biofeedback that have the potential of being extremely specific in their action, modifying one visceral response without altering other closely related visceral responses, have rarely functioned in this way when applied clinically. With a few important exceptions, therapeutic effects associated with biofeedback are accompanied by general relaxation and decreased arousal. Thus, while conceptually biofeedback tends to be viewed as a specific therapy, in practice the more general effects of relaxation may play a disproportionately large role in achieving the therapeutic effects.

CLINICAL APPLICATIONS

Disorders of Cardiac Rate and Rhythm

Numerous studies have shown that normal individuals can be trained to increase or decrease their heart rate reliably, though the degree of change will depend on baseline levels and the individual's overall normal cardiac variability, as well as on the kinds of inducements and other details of the procedure.

This procedure has been used successfully to treat sinus tachycardia, supraventricular tachycardia, and atrial fibrillation by training patients to slow their ventricular heart rate. Pharmacological studies carried out with the latter condition suggest that efferent vagal activity mediates this ability. Further, at least one patient with Wolff-Parkinson-White syndrome was trained to increase or decrease normal AV conduction, again by modifying efferent vagal activity. A group of eight patients with premature ventricular contractions (PVCs) were trained to decrease the number of PVCs by first being taught to increase, decrease, and maintain their heart rate within a

narrow range. Five of the eight patients showed a decrease in PVCs during the course of training (28). Training to control PVCs, as well as other efforts to modify clinical disturbances of cardiac rhythms, requires considerable time. In the cited study, for instance, participation in over 47 training sessions was correlated with improvement. The long-term results of training, however, are variable. One patient showed a reduction of PVC frequency from about ten per minute to approximately one every five minutes by the end of the study, maintaining this improvement for several years. Another patient who reduced her PVCs from twenty per minute to none for approximately fifteen minutes during training did not maintain this improvement at the end of the training period. (For reviews of these studies see 29, 30.) Though the demonstration of learned control over some aspects of cardiac dysfunction is dramatic, other recent work suggests that relaxation training alone may be effective in reducing PVCs (31). While relaxation training does not have the specificity of action seen with biofeedback, the importance of relaxation for clinical improvement, as previously mentioned, deserves attention. In sum, findings to date justify continuing research but do not suggest that biofeedback therapy is ready to be considered as an alternative to accepted drug therapy except on an investigative basis.

Raynaud's Disease

The painful paroxysms of cutaneous vasospasm that characterize Raynaud's disease have been successfully treated with temperature feedback training of the affected limb. Patients with Raynaud's disease can learn to increase finger temperature as readily as normal individuals, and with adequate training are able to raise finger temperature from the occasionally depressed level to the normal range and maintain it there even when challenged by cold pressor tests and other kinds of exposure to cold environments.

A number of individual case reports and series of cases, including one large series of 80 patients, reported excellent results following digital temperature feedback training (48). It would appear important that patients learn to maintain and self-regulate the increase of temperature over a period of time and to retain this ability when feedback training is completed. However, a successful result also depends upon being able to do so without feedback and to maintain the increased finger temperature while carrying on concurrent activities. (For a review, see 32.)

Though the results with Raynaud's disease are extremely encouraging, it should also be recognized that the syndrome can be precipitated by emotional upset and is importantly influenced by psychogenic factors. Further, relaxation training in general and autogenic training in particular, which involves the use of self-suggestive phrases such as "My hands are

heavy and warm," have been used with generally good results, and in one recent study of Raynaud's disease (33) both autogenic training procedures and biofeedback training yielded excellent results that did not differ from each other.

Related vasoconstrictive disorders apparently responsive to biofeedback training include acrocyanosis, scleroderma, pernio syndrome, and trench foot. While the reported results are promising with each of these conditions, further study is needed.

Biofeedback therapy is without any serious side effects and thus would be preferable to either drug therapy or sympathectomy. A therapeutic trial of biofeedback should be initially undertaken and, if effective, should become the treatment of choice for Raynaud's disease.

Migraine

True migraine is characterized by arterial dilation of the cerebral circulation, and there are numerous reports of its successful treatment by training patients to increase finger temperature, thereby inducing a reflex or reciprocal constriction of the cerebral circulation. The reported results range from 50% to over 70% improvement. Cases of classical migraine, with a clear aura and a typical course, respond best (20, 34). There is some controversy, however, whether handwarming is necessarily the specific treatment or is effective only as part of more general relaxation training. Thus, Price & Tursky (35) compared temperature feedback, relaxation training, and a neutral control group; they reported that either type of active treatment was far more effective than the control—supporting the view that learning to increase blood flow of the hands may only reflect a general autonomic relaxation.

Though it is not clear whether handwarming is the specific therapeutic agent or whether it involves the concomitant learning of relaxation, the clinical results reported are positive, and either temperature biofeedback training or other forms of relaxation training are safe and generally effective therapeutic procedures for this ubiquitous and oftentimes extremely troublesome malady.

Tension Headaches

Tension headache is typically accompanied by muscle spasms of the head and neck. Feedback of muscle tension, based on the electromyogram (EMG), typically of the frontalis muscle, has been used with considerable success. A number of studies report lasting results with those patients who not only learned to lower muscle tension but also to carry out relaxation exercises at home (36). Here again it remains to be established whether EMG feedback is uniquely important or whether temperature feedback or

relaxation training may be equally effective. Since most clinicians using biofeedback in the treatment of tension headaches employ a combination of these techniques, this is largely a matter of only academic interest. Clinical results have been sufficiently encouraging to justify a therapeutic trial of biofeedback for this condition.

Hypertension and Hypotension

Following reports that large and persistent changes in blood pressure could be produced in curarized animals using feedback procedures, there was considerable interest in the use of biofeedback for the treatment of high blood pressure. Several studies (37, 38) reported consistent significant decreases during feedback sessions. Typically, a large number of training sessions were required and the decreases were generally not very large. In the Benson study (37) of patients with elevated systolic blood pressures, three of seven patients' blood pressures returned to normal limits, while in the Schwartz & Shapiro study (38) of patients with elevated diastolic pressures, only one of the seven patients showed a clinically meaningful decrease. There are few reports of follow-up, and those that are available indicate that the patients do not tend to maintain the decreased blood pressure. Similar results have also been reported with various kinds of relaxation procedures.

While the treatment of hypertension with presently available biofeedback technology is neither particularly effective nor necessarily specific in its action, dramatic and clinically useful work has been carried out in the treatment of hypotension following spinal cord injuries. Miller and his associates (39, 40), treating paraplegics unable to maintain a vertical position without fainting due to hypotension, have been able to teach a number of such patients sufficient control to maintain satisfactory levels of blood pressure. Though this application is relevant to a very select group of patients, its consequence is nonetheless dramatic and suggests the merit of further research to determine the conditions under which man can learn to regulate his own blood pressure.

Insomnia

Biofeedback in the form of muscle tension training, followed by training to increase theta waves (4–7 cycles per second EEG activity), has been reported to be effective in the treatment of insomnia (41). A number of reports also indicate that relaxation training, in the form of either progressive relaxation, autogenic training, TM, or self-hypnosis, is effective in the treatment of insomnia. Thus, while biofeedback may play a role, it has not yet been documented as more effective than the simpler and more readily available forms of relaxation training. Relaxation training, with or without

biofeedback, is safe, tends to increase the patient's feeling of mastery, and should be considered for the treatment of chronic insomnia (for a review, see 42).

Neuromuscular Rehabilitation

As discussed previously, some of the earliest uses of biofeedback were in the area of rehabilitation. A variety of feedback procedures have been employed with considerable ingenuity and success. There is little doubt that feedback procedures should have a role in the rehabilitation of a wide range of patients with sequelae of trauma, CVAs, and some neurological disorders (for a review, see 43).

TORTICOLLIS This notoriously recalcitrant syndrome has been treated with muscle feedback, relaxation training, and other behavioral procedures with varying success. In a large series, Brudny et al (44-46) used muscle feedback and what the authors call sensory feedback therapy with a group of 40 patients serving as their own controls. Patients were assessed physiologically and evaluated for the degree of head tilt by video tape before and after 8-10 weeks of training. Auditory and visual EMG feedback were used to teach patients to diminish the contraction of the hypertrophied sternocleidomastoid on the ipsilateral side and to increase the contraction of the corresponding atrophied muscle on the contralateral side to achieve postural control. Fifty-six percent of the torticollis patients showed significant clinical improvement and profound positive changes in daily living. Forty percent of the patients maintained improvement at three-months follow-up. Results to date indicate that biofeedback procedures may therefore be useful in the treatment of torticollis; however, for maximal effectiveness, it is likely that they will need to be combined with other behavioral and psychological approaches.

INCONTINENCE One of the most clear-cut and specific uses of biofeedback demonstrated to date is its application to the treatment of fecal incontinence. When the rectum or rectosigmoid colon is stimulated by distention comparable to that produced by fecal matter descending into this area, there is a reflex relaxation of the internal anal sphincter, which normally is compensated for by a brief contraction of the external anal sphincter, preventing the escape of stool at inappropriate times. In many cases of fecal incontinence the external sphincter response is absent or weak, often the sequela of pelvic surgery, trauma, or various neuropathies. A procedure for retraining sphincter control in patients with this type of fecal incontinence was reported by Engel et al (47). Two balloons are used to measure the response of the internal and external sphincters respectively, and a third

balloon is inflated in the rectum to simulate fecal material by producing internal sphincter relaxation and external sphincter contraction. The patient is shown the correct response and his own on a strip recorder and told what he should try to do. The rectal balloon is repeatedly inflated with air and the patient is verbally praised for successfully approximating the desired response. As he improves, the amount of stimulation is progressively decreased until the appropriate external sphincter response is made to the weakest stimulus likely to occur.

The first report included six patients, four of whom became completely continent after treatment, and the other two much improved. Subsequently, an additional 50 patients have been studied, and 72% of these were either completely continent or at least 90% improved after treatment. Two or three training sessions at monthly intervals were usually sufficient for therapeutic results. Considering the surgical alternatives, this is an instance where a safe, simple, and relatively inexpensive procedure can reverse an extremely embarrassing symptom and dramatically enhance the quality of the patient's life.

Though various kinds of general biofeedback therapies have been used for urinary incontinence, no procedure analogous to that for the treatment of fecal incontinence has thus far been applied. It would seem a relatively simple matter to adapt urological procedures to the biofeedback mode, and one might reasonably expect positive results with some types of bladder dysfunction and urinary stress incontinence.

SUMMARY AND CONCLUSIONS

Biofeedback is an exciting new technology, particularly since it helps the patient himself gain mastery over his own difficulty and ultimately to become his own therapist. The information fed back to the patient, usually on a fairly continuous basis, can also be recorded in ways helpful to the physician in assessing the patient's problems. In many circumstances, information that can document gradual improvement is particularly helpful to patients in allowing them to recognize progress which would not otherwise be obvious. This kind of information is likely to be helpful in maintaining the patient's motivation. Further, biofeedback helps the patient take an active role in his own therapy, an aspect especially important in the treatment of chronic conditions.

Biofeedback therapy does not, however, take place in a vacuum, and the nature of the doctor-patient relationship and other psychological factors will profoundly affect the results of biofeedback as well as those of any other therapy. In addition to the information that biofeedback provides by way of an external feedback loop, the various procedures designed to reward the

patient for his performance can, in their own right, play an important therapeutic role. It is also clear that for certain conditions, as in other therapies, some systematic program of follow-up treatment is needed to maintain continuing progress.

While there are some highly promising specific effects of biofeedback—as in the treatment of fecal incontinence—it is also clear that most biofeedback therapies involve training the patient to decrease his level of physiological arousal and therefore serve to bring about the same end results as a wide variety of relaxation therapies. However, the fact that relaxation therapy is often as effective as biofeedback ought not discourage us from the use of biofeedback but rather lead us to recognize the importance of gaining control over those continuing states of hyperarousal that are related to the whole gamut of stress-related symptoms. In a sense, the dramatic technique of biofeedback, employing as it often does the most advanced technology, helps to remind us of the close association between subjective experience and the homeostatic mechanisms controlled by the autonomic nervous system. Further, it has reawakened interest in relaxation training procedures that are perhaps as old as man and can be found in virtually every culture. The techniques of biofeedback have thus not only shown themselves to be therapeutically effective in a number of stress related problems but also have provided the means of both inducing and assessing various forms of mental and physical relaxation, the importance of which has only begun to be rediscovered in modern medicine.

ACKNOWLEDGMENT

I would like to express my appreciation to my colleagues, Peter B. Bloom, David F. Dinges, Frederick J. Evans, Emily Carota Orne, William M. Waid, and Stuart K. Wilson, for their suggestions and critical comments in the preparation of this manuscript. The substantive work carried out in our laboratory relevant to this paper and its preparation was supported in part by grant MH 19156 from the National Institute of Mental Health and by a grant from the Institute for Experimental Psychiatry.

Literature Cited

1. Skinner, B. F. 1935. Two types of conditioned reflex and a pseudo type. *J. Genet. Psychol.* 12:66-77
2. Mowrer, O. H. 1960. *Learning Theory and Behavior*. New York: Wiley. 555 pp.
3. Engel, B. T. 1972. Operant conditioning of cardiac function: A status report. *Psychophysiology* 9:161-77
4. Shapiro, D., Tursky, B., Schwartz, G. E. 1970. Differentiation of heart rate and systolic blood pressure in man by operant conditioning. *Psychosom. Med.* 32:417-23
5. Kemmel, H. D. 1967. Instrumental conditioning of autonomically mediated behavior. *Psychol. Bull.* 67:337-45
6. Shapiro, D., Crider, A. B., Tursky, B. 1964. Differentiation of an autonomic response through operant reinforcement. *Psychonom. Sci.* 1:147-48
7. Basmajian, J. V. 1963. Control and training of individual motor units. *Science* 141:440-41
8. Miller, N. E. 1969. Learning of visceral and glandular responses. *Science* 163:434-45
9. DiCara, L. V., Miller, N. E. 1968. Instrumental learning of vasomotor responses by rats: Learning to respond differentially in the two ears. *Science* 159:1485-86
10. DiCara, L. V. 1970. Learning in the autonomic nervous system. *Sci. Am.* 222:30-39
11. Kamiya, J. 1969. Operant control of the EEG alpha rhythm and some of its reported effects on consciousness. In *Altered States of Consciousness: A Book of Readings*, ed. C. T. Tart, pp. 507-17. New York: Wiley, 575 pp.
12. Brown, B. B. 1971. Awareness of EEG-subjective activity relationships detected within a closed feedback system. *Psychophysiology* 7:451-64
13. Nowlis, D. P., Kamiya, J. 1970. The control of electroencephalographic alpha rhythms through auditory feedback and the associated mental activity. *Psychophysiology* 6:476-84
14. Hart, J. T. 1968. Autoccontrol of EEG alpha. *Psychophysiology* 4:506 (Abstr.)
15. Fox, S. S., Rudell, A. P. 1968. Operant controlled neural event: Formal and systematic approach to electrical coding of behavior in brain. *Science* 162:1299-1302
16. Beatty, J., Greenberg, A., Deibler, W. P., O'Hanlon, J. F. 1974. Operant control of occipital theta rhythm affects performance in a radar monitoring task. *Science* 183:871-73
17. Mulholland, T. 1968. Feedback electroencephalography. *Act. Nerve. Super.* 10:410-38
18. Sterman, M. B., Friar, L. 1972. Suppression of seizures in an epileptic following sensorimotor EEG feedback training. *Electroencephalogr. Clin. Neurophysiol.* 33:89-95
19. Stoyva, J., Budzynski, T. 1974. Cultivated low arousal—an anti-stress response? In *Limbic and Autonomic Nervous System Research*, ed. L. V. DiCara, pp. 369-94. New York: Plenum. 428 pp.
20. Sargent, J. D., Green, E. E., Walters, E. D. 1973. Preliminary report on the use of autogenic feedback techniques in the treatment of migraine and tension headaches. *Psychosom. Med.* 35:129-35
21. Schultz, J. H. 1932. *Das Autogene Training*. Leipzig: Thieme. 410 pp.
22. Jacobson, E. 1938. *Progressive Relaxation*. Chicago: University Press. 494 pp. 3rd ed.
23. Schultz, J. H., Luthe, W. 1959. *Autogenic Training: A Psychophysiological Approach in Psychotherapy*. New York: Grune & Stratton. 289 pp.
24. Luthe, W., ed. 1970. *Autogenic Therapy*. Vols. 1-6. New York: Grune & Stratton
25. Benson, H., Klipper, M. Z. 1975. *The Relaxation Response*. New York: Morrow. 158 pp.
26. Miller, N. E., Dworkin, B. R. 1974. Visceral learning: Recent difficulties with curarized rats and significant problems for human research. In *Cardiovascular Psychophysiology*, ed. P. A. Obrist, A. H. Black, J. Brener, L. V. DiCara, pp. 312-31. Chicago: Aldine. 662 pp.
27. Orne, M. T., Wilson, S. K. 1978. On the nature of alpha feedback training. In *Consciousness and Self-Regulation: Advances in Research and Theory*, Vol. 2, ed. G. E. Schwartz, D. Shapiro, pp. 359-400. New York: Plenum. 445 pp.
28. Weiss, T., Engel, B. T. 1971. Operant conditioning of heart rate in patients with premature ventricular contractions. *Psychosom. Med.* 33:301-21
29. Engel, B. T., Blecker, E. B. 1974. Application of operant conditioning techniques to the control of the cardiac arrhythmias. See Ref. 26, pp. 456-76
30. Weiss, T. 1977. Biofeedback training for cardiovascular dysfunctions. *Med. Clin. North Am.* 61:913-28

31. Benson, H., Alexander, S., Feldman, C. L. 1975. Decreased premature ventricular contractions through the use of the relaxation response in patients with stable ischemic heart-disease. *Lancet* 2: 380-82
32. Taub, E. 1977. Self-regulation of human tissue temperature. In *Biofeedback: Theory and Research*, ed. G. E. Schwartz, J. Beatty, pp. 265-300. New York: Academic. 467 pp.
33. Surwit, R. S., Pilon, R. N., Fenton, C. 1977. *Behavioral treatment of Raynaud's disease*. Presented at Ann. Meet. Assoc. Adv. Behav. Ther., Atlanta, Ga.
34. Blanchard, E. B., Young, L. D. 1974. Clinical application of biofeedback training: A review of evidence. *Arch. Gen. Psychiatry* 30:573-89
35. Price, K. P., Tursky, B. 1976. Vascular reactivity of migraineurs and non-migraineurs: A comparison of responses to self-control procedures. *Headache*, Nov., pp. 210-17
36. Budzynski, T. H., Stoyva, J. M., Adler, C. S., Mullaney, D. J. 1973. EMG biofeedback and tension headache: A controlled outcome study. *Psychosom. Med.* 35:484-96
37. Benson, H., Shapiro, D., Tursky, B., Schwartz, G. E. 1971. Decreased systolic blood pressure through operant conditioning techniques in patients with essential hypertension. *Science* 173: 740-42
38. Schwartz, G. E., Shapiro, D. 1973. Biofeedback and essential hypertension: Current findings and theoretical concerns. In *Biofeedback: Behavioral Medicine*, ed. L. Birk, pp. 133-43. New York: Grune & Stratton. 209 pp.
39. Miller, N. E. 1974. Biofeedback: Evaluation of a new technic. *N. Engl. J. Med.* 290:684-85
40. Miller, N. E. 1976. Fact and fancy about biofeedback and its clinical implications. *Cat. Sel. Doc. Psychol.* 6:92, MS 1329
41. Stoyva, J., Budzynski, T., Sittenfeld, P., Yaroush, R. 1974. A two-step EMG-theta feedback training in sleep onset insomnia: Preliminary results. *Proc. Biofeedback Res. Soc., 5th*, Colorado Springs, p. 103 (Abstr.)
42. Bootzin, R. R., Nicassio, I. M. 1978. Behavioral treatment of insomnia. In *Progress in Behavior Modification*, Vol. 6, ed. M. Hersen, pp. 1-45. New York: Academic
43. Engel-Sittenfeld, P. 1977. Biofeedback in the treatment of neuromuscular disorders. In *Biofeedback and Behavior*, ed. J. Beatty, H. Legewie, pp. 427-48. New York: Plenum. 531 pp.
44. Brudny, J., Grynbaum, B. B., Korein, J. 1973. New therapeutic modality for treatment of spasmodic torticollis. *Arch. Phys. Med.* 54:575 (Abstr.)
45. Brudny, J., Grynbaum, B. B., Friedman, L. W. 1974. Electromyography in the practice of rehabilitation medicine: Therapeutic applications. *Arch. Phys. Med.* 55:564 (Abstr.)
46. Brudny, J., Korein, J., Levidow, L., Grynbaum, B. B., Lieberman, A., Friedman, L. 1974. Sensory feedback therapy as a modality of treatment in central nervous system disorders of voluntary movement. *Neurology* 24: 925-32
47. Engel, B. T., Nikoomanesh, P., Schuster, M. M. 1974. Operant conditioning of rectosphincteric responses in the treatment of fecal incontinence. *N. Engl. J. Med.* 290:646-49
48. Taub, E., Stroebe, C. F. 1978. *The use of biofeedback in the treatment of vasoconstrictive syndromes. 1978. a Task Force Study Section No. 6 report*, pp. 1-16. Denver: Biofeedback Soc. Am.

CUMULATIVE INDEXES

CONTRIBUTING AUTHORS, VOLUMES 26-30

A

Aberdeen, E., 26:451-64
 Abildskov, J. A., 30:171-79
 Ahmed, S. S., 28:393-409
 Ahn, Y. S., 28:299-309
 Alexander, P., 27:207-24
 Alfrey, A. C., 29:93-98
 Annas, G. J., 29:9-14
 Anthony, B. F., 28:355-69
 Arnold, A. B., 28:187-94
 Athanasoulis, C. A., 30:41-59
 Atkinson, J. P., 28:291-98

B

Bagshaw, M. A., 26:567-88
 Bailey, J. C., 29:417-26
 Baldwin, D. S., 27:49-55
 Baldwin, R. W., 27:151-63
 Barchas, J. D., 26:499-520
 Barnes, A. C., 27:1-7
 Begemann, H., 28:131-41
 Beisel, W. R., 26:9-20
 Bergsagel, D. E., 30:431-43
 Berman, M. L., 26:391-401
 Betts, R. F., 28:103-10
 Bierman, E. L., 30:155-70
 Bigger, J. T. Jr., 30:91-118
 Blankenhorn, D. H.,
 30:289-300
 Blum, C. B., 26:365-71
 Bonadonna, G., 29:149-75
 Bone, H. G. III, 28:111-17
 Boshes, B., 26:465-70
 Boyar, R. M., 29:509-20
 Branch, R. A., 29:205-18
 Braverman, L. E., 26:443-49
 Braylan, R. C., 30:17-24
 Brenner, B. M., 28:91-102
 Bresnick, G. H., 30:331-38
 Bristow, J. D., 29:475-83
 Brunner, E. A., 26:391-401
 Burke, J. S., 30:17-24

C

Cadore, R. J., 26:21-25
 Cannon, P. J., 27:485-521
 Cantor, H., 30:269-77
 Caro, W. A., 28:251-69

Cavaliere, R. R., 28:57-65
 Chase, T. N., 26:419-26
 Cheng, S. C., 26:391-401
 Cohen, J. J., 29:521-41
 Cohen, S., 26:373-90
 Colman, R. W., 30:359-74
 Colvin, R. B., 29:191-203
 Conklin, R. H., 28:167-77
 Cooper, T. B., 30:259-68
 Craig, W. A., 27:287-300
 Crawford, D. W., 30:289-300
 Crosnier, J., 29:67-72

D

Dale, D. C., 26:277-81
 D'Angio, G. J., 30:405-15
 Dean, R. H., 27:397-405
 Dement, W. C., 27:465-84
 DeVries, W. C., 26:203-17
 Doherty, J. E., 26:159-71
 Downing, D. T., 26:27-32
 Dudrick, S. J., 28:517-28
 DuPont, H. L., 28:167-77
 Durie, B. G. M., 26:283-88
 Durrer, D., 27:63-68
 Dustan, H. P., 29:485-93

E

Ehrhardt, A. A., 30:417-30
 Eickhoff, T. C., 28:411-23
 Elharrar, V., 29:417-26
 Elin, R. J., 27:127-41
 Elliott, G. R., 26:499-520
 Ensink, J. W., 30:155-70
 Epstein, C. J., 29:117-28
 Ettinger, P. O., 28:393-409
 Ettlinger, R. E., 29:15-22
 Evenson, M. A., 28:85-90
 Everett, R. B., 30:473-88

F

Fajans, S. S., 30:313-29
 Fanelli, G. M. Jr., 28:349-54
 Fang, L. S. T., 30:225-39
 Fauci, A. S., 26:277-81
 Feinstone, S. M., 29:359-66
 Feldman, D., 26:83-90
 Fialkow, P. J., 30:135-43

Fisher, R. S., 26:373-90
 Fitzgerald, F., 29:177-89
 Floyd, J. C. Jr., 30:313-29
 Forget, B. G., 26:345-51
 Forstot, S. L., 28:21-35
 Frank, M. M., 28:291-98
 Freeman, H. J., 29:99-116
 Freeman, R. G., 27:77-87
 Freis, E. D., 30:81-89
 Friedman, R. C., 27:57-62
 Froelicher, V. F. Jr., 28:1-12

G

Gabbay, K. H., 26:521-36
 Gadek, J., 26:291-98
 Gasparini, M., 29:149-75
 Gault, M. H., 27:117-25
 Gavras, H., 27:485-521
 Gennari, F. J., 29:521-41
 Gerald, P. S., 29:563-70
 Gerich, J. E., 27:379-88
 Gibbons, R. J., 26:121-36
 Giebink, G. S., 29:285-306
 Gillis, R. A., 29:73-79
 Glatstein, E., 29:81-91
 Gluck, F. W., 27:397-405
 Golbus, M. S., 29:117-28
 Goldberg, H. I., 30:181-88
 Goldberg, M., 26:537-50
 Goldfine, I. D., 29:407-15
 Goldreyer, B. N., 26:219-28
 Goldsmith, R. S., 27:181-90
 Goodnight, J. E. Jr., 29:231-83
 Graber, S. E., 29:51-66
 Grams, R. R., 27:199-206
 Green, R., 27:57-62
 Greenblatt, D. J., 27:407-20
 Grossman, M., 28:43-55
 Grossman, M. I., 29:495-507
 Grumet, F. C., 28:425-52
 Guillemin, R., 27:379-88
 Guilleminault, C., 27:465-84
 Gunnar, R. M., 26:471-83
 Gunnells, J. C. Jr., 26:259-75
 Gutterman, J. U., 28:489-515

H

Habener, J. F., 29:327-42
 Hackett, T. P., 27:371-78

506 CONTRIBUTING AUTHORS

- Haider, B., 28:393-409
Hall, C. L., 30:199-211
Halimi, K. A., 29:137-48
Hamburger, J., 29:67-72
Hanshaw, J. B., 28:103-10
Hardwicke, J., 30:199-211
Harmel, M. H., 26:137-41
Harrington, W. J., 28:299-309
Hartmann, R. C., 28:187-94
Havel, R. J., 28:195-209
Henderson, E. S., 26:289-306
Henkin, R. E., 26:193-201
Herdon, J. H. Jr., 27:77-87
Hersh, E. M., 28:489-515
Hetzel, D., 30:61-79
Higby, D. J., 26:289-306
Holland, J. F., 28:475-88
Hollister, L. E., 27:431-46
Holman, R. B., 26:499-520
Holmberg, L., 26:33-44
Holmsen, H., 30:119-34
Hook, E. W., 28:371-91
Hopewell, P. C., 27:343-56
Hornick, R. B., 30:457-72
Hoyumpa, A. M. Jr., 29:205-18
Hunder, G. G., 29:15-22
Hurst, M. W., 27:301-12
- I
- Ingbar, S. H., 26:443-49
- J
- Janowitz, H. D., 27:269-85
Jenkins, C. D., 27:301-12; 29:543-62
Johannes, R. S., 26:589-601
Jordan, P. H. Jr., 27:191-97
- K
- Kadir, S., 30:41-59
Kales, A., 30:301-12
Kales, J. D., 30:301-12
Kane, J. J., 26:159-71
Kaplan, B. S., 27:117-25
Kaplan, N. M., 29:31-40
Karnovsky, M. J., 30:213-24
Kattus, A. A., 27:69-75
Kaufman, H. E., 28:21-35
Kaye, M. D., 30:375-404
Kennedy, B. J., 26:235-43
Kim, Y., 29:319-25
Kim, Y. S., 29:99-116
Kinney, D. K., 29:459-73
Kitabchi, A. E., 30:339-57
Kitzmler, J. L., 30:375-404
Klassen, J., 27:117-25
Klein, M. S., 27:89-108
Klein, P. A., 28:311-27
Knochel, J. P., 30:145-53
Korenman, S. G., 27:357-70
Korobkin, M., 30:181-88
Krantz, S. B., 29:51-66
Kunin, C. M., 27:287-300
- L
- Lamberti, J. J., 29:571-82
Langer, G. A., 28:13-20
Lanning, C. F., 26:137-41
Le Quesne, L. P., 26:63-74
Levy, R. I., 26:365-71
Lewis, E. J., 30:445-55
Linnemann, C. C. Jr., 28:179-85
Lipinski, J., 26:551-65
Lipscomb, K., 28:271-89
Littman, P. S., 30:405-15
Lofgren, S., 27:407-20
Long, J. M. III, 28:517-28
Lopez, M., 27:453-63
Love, R., 28:131-41
Lyons, M. M., 28:393-409
- M
- Maas, J. W., 29:219-29
MacDonald, P. C., 30:473-88
MacLean, W. A. H., 30:259-68
Maddox, D. A., 28:91-102
Mahaffey, J. E., 29:327-42
Majerus, P. W., 29:41-49
Malarkey, W. B., 30:249-58
Marchesi, V. T., 29:593-603
Marcus, R., 27:357-70
Matthysse, S., 26:551-65; 29:459-73
Mavligit, G. M., 28:489-515
McCabe, W. R., 27:335-41
McCluskey, R. T., 29:191-203
McDevitt, H. O., 28:425-52
McGraw, C. P., 26:321-29
McGuffin, W. L. Jr., 26:259-75
McGuigan, J. E., 29:307-18
McGuire, W. L., 26:353-63
McPhaul, J. J. Jr., 28:37-42
McRee, D. I., 27:109-15
Meares, E. M. Jr., 30:279-88
Mello, N. K., 27:321-33
Mendelson, J. R., 27:321-33
Merin, R. G., 28:75-83
Merrill, J. P., 29:343-58
Messner, R. P., 26:181-92
Meyer-Bahlburg, H. F. L., 30:417-30
Michael, A. F., 29:319-25
Miletich, J. P., 29:41-49
Millman, B. S., 28:223-34
Minna, J. D., 30:359-74
Modell, J. H., 29:1-8
Moran, J. F., 26:471-83
Morris, L., 30:339-57
Morton, D. L., 29:231-83
Moylan, J. A., 28:85-90
Mullins, C. B., 28:271-89
Murray, J. F., 27:343-56
Murray, T., 26:537-50
Myers, F. L., 30:331-38
- N
- Nasenbeny, J., 28:179-85
Nathan, D. G., 26:345-51
Nicoloff, D. M., 26:143-57
Nilsson, I. M., 26:33-44
Noël, L.-H., 29:67-72
Norman, P. S., 26:337-44
- O
- Oates, J. A., 29:23-29
Odell, W. D., 29:379-406
Okada, D. M., 28:355-69
Oken, D. E., 26:307-19
Oldewurtel, H. A., 28:393-409
Oliver, J. A., 27:485-521
Ontjes, D. A., 27:263-68
Oratz, M., 26:91-104
Orci, L., 28:119-30
O'Reilly, R. A., 27:245-61
Orne, M. T., 30:489-503
Osborn, A. G., 20:189-98
- P
- Pak, C. Y. C., 28:111-17
Parisi, A. F., 28:159-66
Paulson, G. W., 26:75-81
Perloff, M., 28:475-88
Pierce, A. K., 28:211-21
Pitot, H. C., 30:25-39
Plotkin, S. A., 29:583-91
Plumb, V. J., 30:259-68
Pochi, P. E., 26:27-32
Portlock, C. S., 29:81-91
Portnoy, B. L., 28:167-77
Price, M. R., 27:151-63
Prien, E. L. Jr., 26:173-79
Purcell, R. H., 29:359-66
- Q
- Quest, J. A., 29:73-79
Quie, P. G., 29:285-306
Quinn, J. L. III, 26:193-201
- R
- Rackley, C. E., 26:105-20
Raisz, L. G., 29:23-29
Rakowski, H., 27:165-80
Ranganathan, N., 27:165-80
Rapoport, B., 28:57-65
Ray, G. R., 26:567-88
Regan, T. J., 28:393-409
Reiffel, J. A., 30:91-118

Resnekov, L., 29:571-82
 Robboy, S. J., 30:359-74
 Robertson, J., 28:211-21
 Robins, E., 28:67-73
 Rocklin, R. E., 30:375-404
 Rose, R. M., 27:301-12
 Rosenberg, R. D., 29:367-78
 Ross, R., 30:1-15
 Rosse, W. F., 26:331-36
 Rossi, A., 29:149-75
 Rothschild, M. A., 26:91-104
 Rubin, R. H., 30:225-39
 Ruderman, N. B., 26:245-58
 Russell, R. O. Jr., 26:105-20
 Ryan, W. G., 28:143-52

S

Sachar, D. B., 27:269-85
 Sachar, E. J., 27:389-96
 Sacks, H., 30:339-57
 Salmon, S. E., 26:283-88
 Salvaggio, J., 27:453-63
 Sanford, J. P., 26:485-98;
 27:421-29
 Sasahara, A. A., 28:159-66
 Sasazuki, T., 28:425-52
 Schacht, R. G., 27:49-55
 Schenker, S., 29:205-18
 Schiff, G. M., 27:447-52
 Schreiber, S. S., 26:91-104
 Scott, H. W. Jr., 27:397-405
 Sever, J. L., 26:589-601
 Seyberth, H. W., 29:23-29
 Shader, R. I., 27:407-20
 Shah, P. M., 28:235-50
 Shah, S. A., 29:563-70
 Sharma, G. V. R. K.,
 28:159-66
 Shearman, D. J. C., 30:61-79
 Sherlock, S., 27:143-49
 Sherwin, I., 27:37-47

Shoulson, I., 26:419-26
 Shull, H. J., 27:397-405
 Silen, W., 27:9-22
 Silver, M. D., 27:165-80
 Sipes, J. N., 28:371-91
 Skillman, J. J., 27:9-22
 Smith, R. B. III, 26:229-34
 Smith, R. T., 28:311-27
 Snyder, W. H. III, 28:111-17
 Sobel, B. E., 27:89-108
 Soldatos, C. R., 30:301-12
 Soll, A. H., 29:495-507
 Spitzer, R. L., 27:57-62
 Stephen, C. R., 28:153-57
 Stossel, T. P., 29:427-57
 Strauss, J. S., 26:27-32
 Sweeney, D. R., 29:219-29

T

Tancini, G., 29:149-75
 Tarazi, R. C., 29:485-93
 Theml, H., 28:131-41
 Thompson, R. L., 28:371-91
 Tilkian, A., 27:465-84
 Tolkoﬀ-Rubin, N. E.,
 30:225-39
 Tomasi, T. B. Jr., 28:453-65
 Toole, J. F., 26:321-29
 Tow, D. E., 28:159-66
 Twito, D. I., 26:235-43

U

Underwood, L. E., 26:427-41
 Unger, R. H., 28:119-30

V

van Houte, J., 26:131-36
 Van Wyk, J. J., 26:427-41
 Voorhees, J. J., 28:467-73

W

Waldo, A. L., 30:259-68
 Wallace, W. A., 26:1-7
 Walzer, S., 29:563-70
 Ward, L. E., 29:15-22
 Warren, J. W., 30:457-72
 Warren, W. D., 26:229-34
 Waterson, A. P., 27:23-35
 Wechsler, A. S., 28:341-48
 Weiss, H. J., 30:119-34
 Weitzman, E. D., 27:225-43
 Wells, J. L. Jr., 30:259-68
 Wells, S. A. Jr., 27:263-68
 Wender, P. H., 26:45-62
 Wessler, S., 27:313-19
 Wigle, E. D., 27:165-80
 Wiktor, T., 29:583-91
 Wilfert, C. M., 29:129-36
 Williams, R. C. Jr., 26:181-92
 Wilson, D. R., 28:329-39
 Winkelmann, R. K., 28:
 251-59
 Winokur, G., 26:21-25
 Wisneski, J. A., 29:475-83
 Wolfe, W. G., 26:203-17
 Wolff, S. M., 26:277-81;
 27:127-41
 Wolfson, A. R., 29:379-406
 Woodrow, J. C., 26:165-78

Y

Yalow, R. S., 30:241-48
 Yen, S. S. C., 26:403-17
 Young, R., 30:339-57
 Yu, P. N., 26:1-7

Z

Zieve, L., 26:143-57
 Zipes, D. P., 29:417-26

CHAPTER TITLES, VOLUMES 26-30

ALLERGY AND IMMUNOLOGY

Alterations in T- and B-Cells in Human Disease States	R. C. Williams Jr., R. P. Messner	26:181-92
The Antiglobulin Test in Autoimmune Hemolytic Anemia	W. F. Rosse	26:331-36
Immunotherapy (Desensitization) in Allergic Disease	P. S. Norman	26:337-44
IgA-Associated Glomerulonephritis	J. J. McPhaul Jr.	28:37-42
The Current Status of BCG Immunization Against Tuberculosis	T. C. Eickhoff	28:411-23
The Association of Genes in the Major Histocompatibility Complex and Disease Susceptibility	T. Sasazuki, F. C. Grumet, H. O. McDevitt	28:425-52
Structure and Function of Alpha-Fetoprotein	T. B. Tomasi Jr.	28:453-65
BCG as Adjuvant Immunotherapy for Neoplasia	E. M. Hersh, J. U. Gutterman, G. M. Mavligit	28:489-515
Polymyalgia Rheumatica and Giant Cell Arteritis	R. E. Ettlinger, G. C. Hunder, L. E. Ward	29:15-22
Control of the Immune System by Inhibitor and Inducer T Lymphocytes	H. Cantor	30:269-77
Immunobiology of the Maternal-Fetal Relationship	R. E. Rocklin, J. L. Kitzmiller, M. D. Kaye	30:375-404
Immunization Against Typhoid Fever	J. W. Warren, R. B. Hornick	30:457-72

ANESTHESIOLOGY

Ketamine Anesthesia	C. F. Lanning, M. H. Harmel	26:137-41
Effects of Anesthesia on Intermediary Metabolism	E. A. Brunner, S. C. Cheng, M. L. Berman	26:391-401
Effects of Anesthetic Drugs on Myocardial Performance in Man	R. G. Merin	28:75-83
Malignant Hyperpyrexia	C. R. Stephen	28:153-57
Acupuncture: Context and Critique	B. S. Millman	28:223-34
Biology of Drowning	J. H. Modell	29:1-8

CARDIOLOGY

Sudden Death and the Pre-Hospital Phase of Acute Myocardial Infarction	W. A. Wallace, P. N. Yu	26:1-7
Diagnosis and Prevention of Postoperative Deep-Vein Thrombosis	L. P. Le Quesne	26:63-74
Left Ventricular Function in Acute and Chronic Coronary Artery Disease	C. E. Rackley, R. O. Russell Jr.	26:105-20
Clinical Pharmacology of Digitalis Glycosides	J. E. Doherty, J. J. Kane	26:159-71
Mechanisms of Supraventricular Tachycardias	B. N. Goldreyer	26:219-28
Selective Distal Splenorenal Shunt for Bleeding Esophageal Varices	R. B. Smith III, W. D. Warren	26:229-34
Low Renin Hypertension	J. C. Gunnells Jr., W. L. McGuffin Jr.	26:259-75

The Steal Syndromes	J. F. Toole, C. P. McGraw	26:321-29
The Total Correction of Congenital Heart Disease	E. Aberdeen	26:451-64
Pathophysiology of Heart Block	J. F. Moran, R. M. Gunnar	26:471-83
Medical and Surgical Treatment of the Wolff-Parkinson-White Syndrome	D. Durrer	27:63-68
Coronary Artery Spasm: A New Appraisal	A. A. Kattus	27:69-75
Medical Management of Myocardial Infarction	M. S. Klein, B. E. Sobel	27:89-108
Mitral Valve Prolapse	E. D. Wigle, H. Rakowski, N. Ranganathan, M. D. Silver	27:165-80
The Detection of Asymptomatic Coronary Artery Disease	V. F. Froelicher Jr.	28:1-12
Ionic Basis of Myocardial Contractility	G. A. Langer	28:13-20
Hypertrophic Cardiomyopathy	P. M. Shah	28:235-50
Critique of Coronary Artery Bypass Surgery	C. B. Mullins, K. Lipscomb	28:271-89
Development of Coronary Collateral Circulation	A. S. Wechsler	28:341-48
Prophylaxis of Infective Endocarditis: A Reevaluation	J. N. Sipes, R. L. Thompson, E. W. Hook	28:371-91
The Role of Ethanol in Cardiac Disease	T. J. Regan, P. O. Ettinger, B. Haider, S. S. Ahmed, H. A. Oldewurtel, M. M. Lyons	28:393-409
Slow-Channel Depolarization: Mechanism and Control of Arrhythmias	J. C. Bailey, V. Elharrar, D. P. Zipes	29:417-26
Left Ventricular Stiffness	J. A. Wisneski, J. D. Bristow	29:475-83
Cardiogenic Hypertension	H. P. Dustan, R. C. Tarazi	29:485-93
Cardiac Assist Devices	J. J. Lamberti, L. Resnekov	29:571-82
The Arterial Wall and Atherosclerosis	R. Ross	30:1-15
Validity of Therapy for Mild Hypertension	E. D. Freis	30:81-89
Sick Sinus Syndrome	J. T. Bigger Jr., J. A. Reiffel	30:91-118
The Prolonged QT Interval	J. A. Abildskov	30:171-79
Studies of Atrial Flutter Following Open Heart Surgery	A. L. Waldo, J. L. Wells Jr., V. J. Plumb, T. B. Cooper, W. A. H. MacLean	30:259-68
Regression of Atherosclerosis	D. W. Crawford, D. H. Blankenhorn	30:289-300
CLINICAL PHARMACOLOGY		
Tardive Dyskinesia	G. W. Paulson	26:75-81
Clinical Pharmacology of Digitalis Glycosides	J. E. Doherty, J. J. Kane	26:159-71
Analgesic Abuse and Renal Disease	T. Murray, M. Goldberg	26:537-50
Rational Psychopharmacology for Patients with Medical Diseases	D. J. Greenblatt, R. I. Shader, S. Lofgren	27:407-20
Effective Use of Analgesic Drugs	L. E. Hollister	27:431-46
Neural Actions of Digitalis	R. A. Gillis, J. A. Quest	29:73-79
The Disposition and Effects of Sedatives and Analgesics in Liver Disease	A. M. Hoyumpa Jr., R. A. Branch, S. Schenker	29:205-18
DERMATOLOGY		
Skin Lipids and Acne	J. S. Strauss, P. E. Pochi, D. T. Downing	26:27-32
Human Disease Associated with Exposure to Light	J. H. Herndon Jr., R. G. Freeman	27:77-87
Current Problems in Mycosis Fungoides and Sezary Syndrome	R. K. Winkelmann, W. A. Caro	28:251-69
Pathophysiology of Psoriasis	J. J. Voorhees	28:467-73

510 CHAPTER TITLES

ENDOCRINOLOGY AND METABOLISM

Metabolic Response to Infection	W. R. Beisel	26:9-20
The Role of Hormone Receptors in the Action of Adrenal Steroids	D. Feldman	26:83-90
Regulation of Albumin Metabolism	M. A. Rothschild, M. Oratz, S. S. Schreiber	26:91-103
Muscle Amino Acid Metabolism and Gluconeogenesis	N. B. Ruderman	26:245-58
Low Renin Hypertension	J. C. Gunnells Jr., W. L. McGuffin Jr.	26:259-75
Endocrine Therapy of Breast Cancer	W. L. McGuire	26:353-63
Interconversions of Apolipoprotein Fragments	C. B. Blum, R. I. Levy	26:365-71
Effects of Anesthesia on Intermediary Metabolism	E. A. Brunner, S. C. Cheng, M. L. Berman	26:391-401
Gonadotropin-Releasing Hormone	S. S. C. Yen	26:403-17
Relation Between Growth Hormone and Somatomedin	J. J. Van Wyk, L. E. Underwood	26:427-41
Active Form of the Thyroid Hormone and Somatomedin	S. H. Ingbar, L. E. Braverman	26:443-49
Hyperglycemia, Polyol Metabolism, and the Complications of Diabetes Mellitus	K. H. Gabbay	26:521-36
Potential Microwave Injuries in Clinical Medicine	D. I. McRee	27:109-15
The Effects of Calcium and Phosphorus in Hemodialysis	R. S. Goldsmith	27:181-90
Circadian Rhythms and Episodic Hormone Secretion in Man	E. D. Weitzman	27:225-43
Estrogens and the Human Male	R. Marcus, S. G. Korenman	27:357-70
Somatostatin: Physiological and Clinical Significance	R. Guillemin, J. E. Gerich	27:379-88
Medical Aspects of Riot Control (Harassing) Agents	J. P. Sanford	27:421-29
Impaired Peripheral Conversion of Thyroxine to Triiodothyronine	R. R. Cavalieri, B. Rapoport	28:57-65
Diagnosis of Hyperparathyroidism	H. C. Bone III, W. H. Snyder, C. Y. C. Pak	28:111-17
The Role of Glucagon in the Endogenous Hyperglycemia of Diabetes Mellitus	R. H. Unger, L. Orci	28:119-30
Paget's Disease of Bone	W. G. Ryan	28:143-52
Classification of the Hyperlipidemias	R. J. Havel	28:195-209
Prostaglandins and Hypercalcemic States	H. W. Seyberth, L. G. Raisz, J. A. Oates	29:23-29
Clinical Hypophosphatemia	F. Fitzgerald	29:177-89
Osteomalacia and Disorders of Vitamin D Metabolism	J. F. Habener, J. E. Mahaffey	29:327-42
Humoral Syndromes Associated with Cancer	W. D. Odell, A. R. Wolfson	29:379-406
Medical Treatment of Acromegaly	I. D. Goldfine	29:407-15
Control of the Onset of Puberty	R. M. Boyar	29:509-20
The Syndrome of Hyporeninemic Hypoadosteronism	J. P. Knoche	30:145-53
Dietary Management of Diabetes Mellitus	J. W. Ensink, E. L. Bierman	30:155-70
Big ACTH and Bronchogenic Carcinoma	R. S. Yalow	30:241-48
Prolactin and the Diagnosis of Pituitary Tumors	W. B. Malarkey	30:249-58
Diagnosis and Medical Management of Insulinomas	S. S. Fajans, J. C. Floyd Jr.	30:313-29
Vitrectomy Surgery for Diabetic Retinopathy	G. H. Bresnick, F. L. Myers	30:331-38
Diabetic Ketoacidosis: Reappraisal of Therapeutic Approach	A. E. Kitabchi, R. Young, H. Sacks, L. Morris	30:339-57

GASTROINTESTINAL TRACT

Regulation in Albumin Metabolism	M. A. Rothschild, M. Oratz, S. S. Schreiber	26:91-104
Pathogenesis of Hepatic Coma	L. Zieve, D. M. Nicoloff	26:143-57
Selective Distal Splenorenal Shunt for Bleeding Esophageal Varices	R. B. Smith III, W. D. Warren	26:229-34
Disorders of the Lower Esophageal Sphincter	R. S. Fisher, S. Cohen	26:373-90
Stress Ulceration in the Acutely Ill	J. J. Skillman, W. Silen	27:9-22
Infectious Particles in Hepatitis	A. P. Waterson	27:23-35
Parietal Cell Vagotomy Without Drainage for the Treatment of Duodenal Ulcer	P. H. Jordan Jr.	27:191-97
New Observations in Crohn's Disease	H. D. Janowitz, D. B. Sachar	27:269-85
Metabolic Complications of Jejunoileal Bypass Operations for Morbid Obesity	H. W. Scott Jr., R. H. Dean, H. J. Shull, F. W. Gluck	27:397-405
Viral Agents and Diarrheal Illness	H. L. DuPont, L. Portnoy, R. H. Conklin	28:167-77
Digestion and Absorption of Protein	H. J. Freeman, Y. S. Kim	29:99-116
Gastrointestinal Hormones	J. E. McGuigan	29:307-18
Non-A, Non-B Hepatitis	S. M. Feinstone, R. H. Purcell	29:359-66
Cellular Mechanisms in Acid Secretion	A. H. Soll, M. I. Grossman	29:495-507
Angiographic Management of Gastrointestinal Bleeding	S. Kadir, C. A. Athanasoulis	30:41-59
The Medical Management of Peptic Ulcer	D. J. C. Shearman, D. Hetzel	30:61-79

HEMATOLOGY

Von Willebrand's Disease	L. Holmberg, I. M. Nilsson	26:33-44
Diagnosis and Prevention of Postoperative Deep-Vein Thrombosis	L. P. Le Quesne	26:63-74
Cellular Kinetics, Staging, and Immunoglobulin Synthesis in Multiple Myeloma	B. G. M. Durie, S. E. Salmon	26:283-88
Granulocyte Transfusion Therapy	D. J. Higby, E. S. Henderson	26:289-306
The Antiglobulin Test in Autoimmune Hemolytic Anemia	W. F. Rosse	26:331-36
Thalassemia	B. G. Forget, D. G. Nathan	26:345-51
Hemochromatosis: Course and Treatment	S. Sherlock	27:143-49
Vitamin K and the Oral Anticoagulant Drugs	R. A. O'Reilly	27:245-61
Medical Management of Venous Thrombosis	S. Wessler	27:313-19
Factors in the Pathomechanism of Chronic Lymphocytic Leukemia	H. Thiel, R. Love, H. Begemann	28:131-41
Paroxysmal Nocturnal Hemoglobinuria (PNH) as a Clonal Disorder	R. C. Hartmann, A. B. Arnold	28:187-94
Cold Agglutinins and Cold-Agglutinin Disease	M. M. Frank, J. P. Atkinson, J. Gadek	28:291-98
Treatment of Idiopathic Thrombocytopenic Purpura (ITP)	Y. S. Ahn, W. J. Harrington	28:299-309
Relationships Between Platelets and Coagulation Factors in Hemostasis	P. W. Majerus, J. P. Miletich	29:41-49
Erythropoietin and the Control of Red Cell Production	S. E. Graber, S. B. Krantz	29:51-66
Heparin, Antithrombin, and Abnormal Clotting	R. D. Rosenberg	29:367-78
Contractile Proteins in Cell Structure and Function	T. P. Stossel	29:427-57
Hairy Cell Leukemia	R. C. Braylan, J. S. Burke	30:17-24
Secretable Storage Pools in Platelets	H. Holmsen, H. J. Weiss	30:119-34
Disseminated Intravascular Coagulation: A Reappraisal	R. W. Colman, S. J. Robboy, J. D. Minna	30:359-74
Treatment of Plasma Cell Myeloma	D. E. Bergsagel	30:431-43

512 CHAPTER TITLES

INFECTION

Metabolic Response to Infection	W. R. Beisel	26:9-20
Granulocyte Transfusion Therapy	D. J. Higby, E. S. Henderson	26:289-306
Urinary Tract Symptoms and Infections	J. P. Sanford	26:485-98
Biology of Endotoxin	R. J. Elin, S. M. Wolff	27:127-40
Significance of Serum Protein and Tissue Binding of Antimicrobial Agents	W. A. Craig, C. M. Kunin	27:287-300
Immunoprophylaxis of Gram-Negative Bacillary Infections	W. R. McCabe	27:335-41
Reye's Syndrome	G. M. Schiff	27:447-52
Cytomegalovirus (CMV) in the Compromised Host(s)	R. F. Betts, J. B. Hanshaw	28:103-10
Pertussis in the Adult	C. C. Linnemann Jr., J. Nasenbeny	28:179-85
The Emergence of Group B Streptococci in Infections of the Newborn Infant	B. F. Anthony, D. M. Okada	28:355-69
E. Coli Meningitis: K ₁ Antigen and Virulence	C. M. Wilfert	29:129-36
Otitis Media: The Spectrum of Middle Ear Inflammation	G. S. Geibink, P. G. Quie	29:285-306
Chronic Bacteremia and Nephritis	Y. Kim, A. F. Michael	29:319-25
Rabies Vaccination	S. A. Plotkin, T. Wiktor	29:583-91
Prostatitis	E. M. Meares, Jr.	30:279-88

KIDNEYS AND URINARY TRACT

Calcium Oxalate Renal Stones	E. L. Prien Jr.	26:173-79
Low Renin Hypertension	J. C. Gunnells Jr., W. L. McGuffin Jr.	26:259-75
Acute Renal Failure (Vasomotor Nephropathy): Micropuncture Studies of the Pathogenetic Mechanism	D. E. Oken	26:307-19
Urinary Tract Symptoms and Infections	J. P. Sanford	26:485-98
Analgesic Abuse and Renal Disease	T. Murray, M. Goldberg	26:537-50
Late Sequelae of Poststreptococcal Glomerulonephritis	D. S. Baldwin, R. G. Schacht	27:49-55
Glomerular Injury in Patients with Neoplasia	B. S. Kaplan, J. Klassen, M. H. Gault	27:117-25
Interrelations of Renin, Angiotensin II, and Sodium in Hypertension and Renal Failure	H. Gavras, J. A. Oliver, P. J. Cannon	27:485-521
Glomerular Filtration of Fluid and Macromolecules: The Renal Response to Injury	D. A. Maddox, B. M. Brenner	28:91-102
Renal Function During and Following Obstruction	D. R. Wilson	28:329-39
Urate Excretion	G. M. Fanelli Jr.	28:349-54
Recurrent Glomerulonephritis After Renal Transplantation	J. Hamburger, J. Crosnier, L.-H. Noël	29:67-72
Dialysis Encephalopathy Syndrome	A. C. Alfrey	29:93-98
Immunological Aspects of Renal Tubular and Interstitial Diseases	R. T. McCluskey, R. B. Colvin	29:191-203
Dialysis Versus Transplantation in the Treatment of End-Stage Renal Disease	J. P. Merrill	29:343-58
Renal Tubular Acidosis	F. J. Gennari, J. J. Cohen	29:521-41
Low Molecular Weight Proteinuria	C. L. Hall, J. Hardwicke	30:199-211
The Ultrastructure of Glomerular Filtration	M. J. Karnovsky	30:213-24
Localization and Antibiotic Management of Urinary Tract Infection	L. S. T. Fang, N. E. Tolckoff-Rubin, R. H. Rubin	30:225-39
Patterns of Circulating Complement in Renal Diseases	E. J. Lewis	30:445-55

MISCELLANEOUS

Dental Caries	R. J. Gibbons, J. van Houte	26:121-36
Unusual Etiologies of Fever and Their Evaluation	S. M. Wolff, A. S. Fauci, D. C. Dale	26:277-81

Implications of Mass Automated Instruments on Medical Practice	R. R. Grams	27:199-206
Confidentiality in Medical Practice	M. Grossman	28:43-55
Informed Consent	G. J. Annas	29:9-14
Recent Membrane Research and Its Implications for Clinical Medicine	V. T. Marchesi	29:593-603
NEOPLASIA		
Scanning Techniques to Assess Thyroid Nodules	J. L. Quinn III, R. E. Henkin	26:193-201
Treatment of Testicular Cancer	D. I. Twito, B. J. Kennedy	26:235-43
Cellular Kinetics, Staging, and Immunoglobulin Synthesis in Multiple Myeloma	B. G. M. Durie, S. E. Salmon	26:283-88
Endocrine Therapy of Breast Cancer	W. L. McGuire	26:353-63
The Role of Radiation Therapy in the Definitive Treatment of Adenocarcinoma of the Prostate	G. R. Ray, M. A. Bagshaw	26:567-88
Tumor Antigens and Tumor-Host Relationships	R. W. Baldwin, M. R. Price	27:151-63
The Functions of the Macrophage in Malignant Disease	P. Alexander	27:207-24
Multiple Endocrine Neoplasia Type II	S. A. Wells Jr., D. A. Ontjes	27:263-68
The Role of Oncogenic Viruses in Neoplasia	P. A. Klein, R. T. Smith	28:311-27
Surgical Adjuvant Chemotherapy	M. Perloff, J. F. Holland	28:475-88
The Non-Hodgkin's Lymphomas: Current Concepts and Management	C. S. Porlock, E. Glatstein	29:81-91
Chemotherapy in Prevention of the Recurrence of Resectable Cancer	G. Bonadonna, G. Tancini, A. Rossi, M. Gasparini	29:149-75
Immunotherapy for Malignant Disease	J. E. Goodnight Jr., D. L. Morton	29:231-83
Biological and Enzymatic Events in Chemical Carcinogenesis	H. C. Pitot	30:25-39
Clonal Origin of Human Tumors	P. J. Fialkow	30:135-43
NEUROLOGY		
The Minimal Brain Dysfunction Syndrome	P. H. Wender	26:45-62
Tardive Dyskinesia	G. W. Paulson	26:75-81
Pathogenesis of Hepatic Coma	L. Zieve, D. M. Nicoloff	26:143-57
The Steal Syndromes	J. F. Toole, C. P. McGraw	26:321-29
Huntington's Disease	I. Shoulson, T. N. Chase	26:419-26
A Definition of Cerebral Death	B. Boshes	26:465-70
Neuroregulators and Sleep Mechanisms	R. B. Holman, G. R. Elliott, J. D. Barchas	26:499-520
Subacute Sclerosing Panencephalitis	R. S. Johannes, J. L. Sever	26:589-601
Temporal Lobe Epilepsy: Neurological and Behavioral Aspects	I. Sherwin	27:37-47
NUTRITION		
Applications and Hazards of Total Intravenous Hyperalimentation	S. J. Dudrick, J. M. Long III	28:517-28
OPHTHALMOLOGY AND OTORHINOLARYNGOLOGY		
Corneal Transplantation	S. L. Forstot, H. E. Kaufman	28:21-35
PSYCHIATRY		
X-Linkage in Manic-Depressive Illness	R. J. Cadoret, G. Winokur	26:21-25
Biochemical Aspects of Schizophrenia	S. Matthyse, J. Lipinski	26:551-65
Reassessment of Homosexuality and Transsexualism	R. C. Friedman, R. Green, R. L. Spitzer	27:57-62
The Relation of Psychological Stress to Onset of Medical Illness	M. W. Hurst, C. D. Jenkins, R. M. Rose	27:301-12
Behavioral and Biochemical Interrelations in Alcoholism	J. H. Mendelson, N. K. Mello	27:321-33

514 CHAPTER TITLES

Psychological Assistance for the Dying Patient and His Family	T. P. Hackett	27:371-78
Neuroendocrine Dysfunction in Depressive Illness	E. J. Sachar	27:389-96
The Sleep Apnea Syndromes	C. Guilleminault, A. Tilkian, W. C. Dement	27:465-84
New Concepts in the Diagnosis of Psychiatric Disorders	E. Robins	28:67-73
Anorexia Nervosa: Recent Investigations	K. A. Halmi	29:137-48
Specificity of Depressive Diseases	D. R. Sweeney, J. W. Maas	29:219-29
Genetic Transmission of Schizophrenia	D. K. Kinney, S. Matthysse	29:459-73
Behavioral Risk Factors in Coronary Artery Disease	C. D. Jenkins	29:543-62
The XYY Genotype	S. Walzer, P. S. Gerald, S. A. Shah	29:563-70
Management of Insomnia	C. R. Soldatos, A. Kales, J. D. Kales	30:301-12
Prenatal Sex Hormones and the Developing Brain: Effects on Psychosexual Differentiation and Cognitive Function	A. A. Ehrhardt, H. F. L. Meyer-Bahlburg	30:417-30
The Efficacy of Biofeedback Therapy	M. T. Orne	30:489-503
RADIOLOGY		
Diagnosis and Prevention of Postoperative Deep-Vein Thrombosis	L. P. Le Quesne	26:63-74
Scanning Techniques to Assess Thyroid Nodules	J. L. Quinn III, R. E. Henkin	26:193-201
The Role of Radiation Therapy in the Definitive Treatment of Adenocarcinoma of the Prostate	G. R. Ray, M. A. Bagshaw	26:567-88
Computed Tomography of the Hepatobiliary System	M. Korobkin, H. I. Goldberg	30:181-88
Computed Tomography in Neurologic Diagnosis	A. G. Osborn	30:189-98
Growth Considerations in the Radiation Therapy of Children with Cancer	P. S. Littman, G. J. D'Angio	30:405-15
REPRODUCTION		
Gonadotropin-Releasing Hormone	S. S. C. Yen	26:403-17
Some Implications of Population Control	A. C. Barnes	27:1-7
Cardiovascular Complications of Oral Contraceptives	N. M. Kaplan	29:31-40
The Prenatal Diagnosis of Genetic Disorders	C. J. Epstein, M. S. Golbus	29:117-28
Endocrinology of the Placenta	R. B. Everett, P. C. MacDonald	30:473-88
RESPIRATORY SYSTEM		
Oxygen Toxicity	W. G. Wolfe, W. C. DeVries	26:203-17
The Adult Respiratory Distress Syndrome	P. C. Hopewell, J. F. Murray	27:343-56
Hypersensitivity Pneumonitis: Current Concepts of Etiology and Pathogenesis	M. Lopez, J. Salvaggio	27:453-63
Diagnosis and Treatment of Fat Embolism	J. A. Moylan, M. A. Evenson	28:85-90
Diagnosis of Pulmonary Embolism	G. V. R. K. Sharma, D. E. Tow, A. F. Parisi, A. A. Sasahara	28:159-66
Pulmonary Complications of General Surgery	A. K. Pierce, J. Robertson	28:211-21